

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE INFORMÁTICA

Departamento de Arquitectura de Computadores y Automática

**MÁSTER EN INVESTIGACIÓN EN INFORMÁTICA
2009-2010**



**TÉCNICAS DE PROCESAMIENTO DE
IMÁGENES ESTEREOSCÓPICAS**

**PROYECTO FIN DE MÁSTER EN INGENIERÍA INFORMÁTICA
PARA LA INDUSTRIA**

ALUMNO:

José Miguel Guerrero Hernández

DIRECTORES:

Gonzalo Pajares Martinsanz
María Guijarro Mata-García

Madrid, 2010

AUTORIZACIÓN

El abajo firmante, matriculado en el Máster en Investigación en Informática de la Facultad de Informática, autoriza a la Universidad Complutense de Madrid (UCM) a difundir y utilizar con fines académicos, no comerciales y mencionando expresamente a su autor el presente Trabajo Fin de Máster: “Técnicas de procesamiento de imágenes estereoscópicas”, realizado durante el curso académico 2009-2010 bajo la dirección de Gonzalo Pajares Martinsanz y María Guijarro Mata-García en el Departamento de Arquitectura de Computadores y Automática, y a la Biblioteca de la UCM a depositarlo en el Archivo Institucional E-Prints Complutense con el objeto de incrementar la difusión, uso e impacto del trabajo en Internet y garantizar su preservación y acceso a largo plazo.

Fdo.: José Miguel Guerrero Hernández

A mis padres, María Rosa y José Miguel.

A Verónica.

ÍNDICE

AUTORIZACIÓN.....	I
RESUMEN Y ABSTRACT	VII
1. INTRODUCCIÓN.....	1
1.1. VISION ESTEREOSCÓPICA	1
1.2. GEOMETRÍA DEL SISTEMA ESTEREOSCÓPICO.....	3
1.3. OBTENCIÓN DE LA DISTANCIA	4
1.4. MOTIVACIÓN DE LA INVESTIGACIÓN.....	5
1.5. PLANTEAMIENTO DEL PROBLEMA	6
1.5.1. Objetivos generales	6
1.5.2. Objetivos específicos.....	7
1.6. APORTACIONES A LA INVESTIGACIÓN	7
1.7. ORGANIZACIÓN DE LA MEMORIA	8
2. REVISIÓN DE MÉTODOS	9
2.1. VISION GENERAL	9
2.2. MÉTODOS BASADOS EN EL ÁREA	11
2.3. MÉTODOS BASADOS EN CARACTERÍSTICAS	14
3. PROPUESTA DE MÉTODOS.....	17
3.1. INTRODUCCIÓN Y MEDIDAS DEL ERROR	17
3.2. CORRELACIÓN DE PEARSON	18
3.3. MINIMIZACIÓN DEL ERROR GLOBAL DE ENERGÍA.....	22
3.4. CORRESPONDENCIA ESTEREOSCÓPICA BASADA EN LÍNEA DE CRECIMIENTO	23
3.5. CORRESPONDENCIA ESTEREOSCÓPICA BASADA EN LA SEGMENTACIÓN	25
3.6. FILTRADO DE IMÁGENES	29
3.6.1. Media.....	29
3.6.2. Mediana.....	30
3.6.3. Enfriamiento simulado (Simulated Annealing).....	32
3.6.4. Propagación.....	35
4. RESULTADOS.....	39
4.1. CONSIDERACIONES PRELIMINARES	39
4.2. ANÁLISIS DE LOS ALGORITMOS PARA LA OBTENCIÓN DE LA DISPARIDAD	41
4.2.1. Correlación de Pearson.....	41
4.2.2. Minimización del error global de energía.....	43
4.2.3. Correspondencia estereoscópica basada en línea de crecimiento.....	45
4.2.4. Correspondencia estereoscópica basada en la segmentación	48
4.3. ANÁLISIS DE LOS FILTROS.....	52
4.3.1. Filtro de la media.....	52
4.3.2. Filtro de la mediana.....	56
4.3.3. Enfriamiento simulado (Simulated Annealing).....	59
4.3.4. Propagación	61
4.4. RESUMEN DE LOS RESULTADOS	63

5. CONCLUSIONES Y TRABAJO FUTURO.....	67
5.1. CONCLUSIONES	67
5.2. TRABAJO FUTURO	68
6. BIBLIOGRAFÍA	73

RESUMEN

En este trabajo se presenta una revisión de métodos para correspondencia estereoscópica con el fin de tratar de encontrar la mejor de las posibles además de tratar de mejorarla. Existen diversos métodos en la literatura para llevar a cabo la correspondencia estereoscópica, siendo generalmente necesario el diseño de nuevos métodos con el fin de mejorar los resultados cuando las técnicas se aplican a entornos reales de exterior. Este es el objetivo del presente trabajo con el fin de aplicar los métodos a entornos reales. Se aprovechan las bondades de los algoritmos estudiados para la realización del nuevo método buscado. Se utiliza técnicas de filtrado comunes tales como la media o la mediana junto con un método de enfriamiento simulado (simulated annealing) y otro de propagación que se ha creado como aporte de investigación. Los resultados de estas dos técnicas se comparan con los métodos clásicos previamente seleccionados.

Palabras clave: visión estereoscópica, correspondencia, filtro de la media, filtro de la mediana, propagación, enfriamiento simulado.

ABSTRACT

This work presents a revision of classical methods in stereovision matching. This is intended for selecting the best approach and to improve its performance. There are several methods for conducting stereo vision matching in the literature, being usually necessary to design new methods to be applied in real outdoor environments. This is the main goal of this work. This is the objective of this work, which presents different techniques for stereo vision, which takes advantage of the benefits of the algorithms used to perform a new method to improve outcome. It uses common filtering techniques such as mean or median along a simulated annealing technique and a spread that has been created as a research contribution. The results of these two techniques are compared with classical methods which are previously selected.

Key words: stereoscopic vision, average filtering, median filtering, propagation, simulated annealing.

INTRODUCCIÓN

1.1. Visión estereoscópica

En los últimos años ha crecido sustancialmente el desarrollo de aplicaciones enfocadas a la tecnología de Visión por Computador, cuyo objetivo principal es proporcionar al sistema de las sensaciones realistas del mundo que se le presenta. Uno de los conceptos que caracteriza la Visión por Computador es la tridimensionalidad; la sensación tridimensional se consigue mediante la obtención de imágenes estereoscópicas generadas por dos cámaras, las cuales adquieren dos imágenes distintas de la misma escena. Es por ello que la obtención de la distancia es la principal preocupación de los sistemas diseñados para la navegación robótica, donde el robot necesita conocer en cada momento la estructura de la escena para navegar en el entorno.

La visión estereoscópica constituye un procedimiento para la obtención de la forma de los objetos en la escena. En este caso la forma se determina a través de la distancia de los objetos en relación con un sistema de referencia por lo que se trata de un método para la obtención de la tercera dimensión.

Para la obtención de la estructura de la escena, existen dos tipos de métodos: *activos* y *pasivos*. Los métodos pasivos son aquellos que intervienen externamente sobre la escena, bien iluminándola o bien enviando un haz energético, tales como sensores de ultrasonidos, luz estructurada, triangulación, telémetro de tiempo de vuelo, etc. Por el contrario, los pasivos no actúan sobre la escena, siendo la *visión estereoscópica*, el principal método que constituye el objetivo final de estudio en este trabajo.

La visión estereoscópica toma como referencia el modelo estereoscópico biológico donde el desplazamiento relativo de los ojos permite obtener la profundidad de los objetos o tercera dimensión mediante un simple proceso de triangulación a partir de las dos imágenes generadas por el mismo objeto de la escena 3D en cada ojo. Esto se debe al hecho de que los ojos están distanciados, esto hace que las imágenes de los objetos en sendos ojos se muestren desplazadas según la distancia de los objetos a los ojos. Si se solapan las imágenes obtenidas en ambos ojos se obtiene la imagen de la

Figura 1.1(b) en la que se observa que la separación relativa entre las imágenes de los dos triángulos es menor que la separación relativa entre las imágenes de las estrellas. Este fenómeno se explica por el hecho de que la estrella en la escena 3-D se encuentra más próxima a los ojos que el triángulo. Estas separaciones relativas de los objetos en las imágenes obtenidas en cada ojo, es lo que se denomina *disparidad* y que es el centro de estudio de este trabajo.

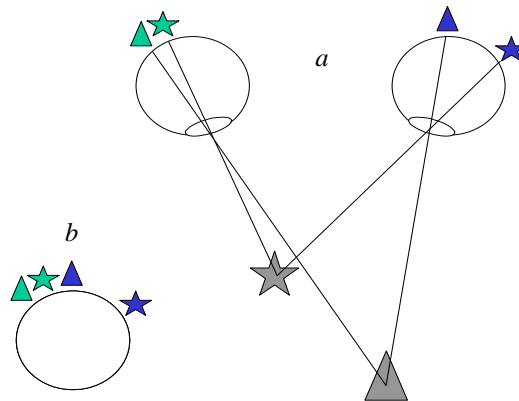


Figura 1.1 - (a) Sistema de visión estereoscópica biológico; (b) Superposición de las imágenes de ambos ojos, donde se observa la separación relativa de cada objeto de la escena (a menor distancia, más lejos se encuentra el objeto; a menor distancia, más cerca está).

En visión estereoscópica artificial generalmente se utilizan dos cámaras separadas entre sí una cierta distancia relativa con las que se obtienen las correspondientes imágenes del par estéreo. El procedimiento consiste en captar dos imágenes de una misma escena, cada imagen es capturada desde una posición de las cámaras ligeramente diferente, por lo que las imágenes se presentan también ligeramente desplazadas entre sí, siendo éste el fundamento básico de la visión estereoscópica, ya que este hecho es el que va a permitir la obtención de la distancia a la que se encuentra un determinado objeto.

En la Figura 1.2 (a) y (b) se muestra un par de imágenes estereoscópicas originales captadas mediante un sistema de visión artificial con dos cámaras alineadas horizontalmente, de forma que los objetos en las imágenes sólo presentan un desplazamiento horizontal y no vertical.

La captura de las imágenes de la escena se puede obtener por alguno de los dos procedimientos siguientes:

- a) Alineando dos cámaras de forma que se sitúen ligeramente desplazadas en el espacio.

- b) Desplazando una única cámara una cierta distancia y captando las imágenes en las diferentes posiciones de desplazamiento.

En ambos casos la geometría del sistema puede diseñarse de forma que las cámaras tengan sus ejes ópticos paralelos o convergentes. El modelo más utilizado en visión artificial es el de ejes ópticos paralelos.

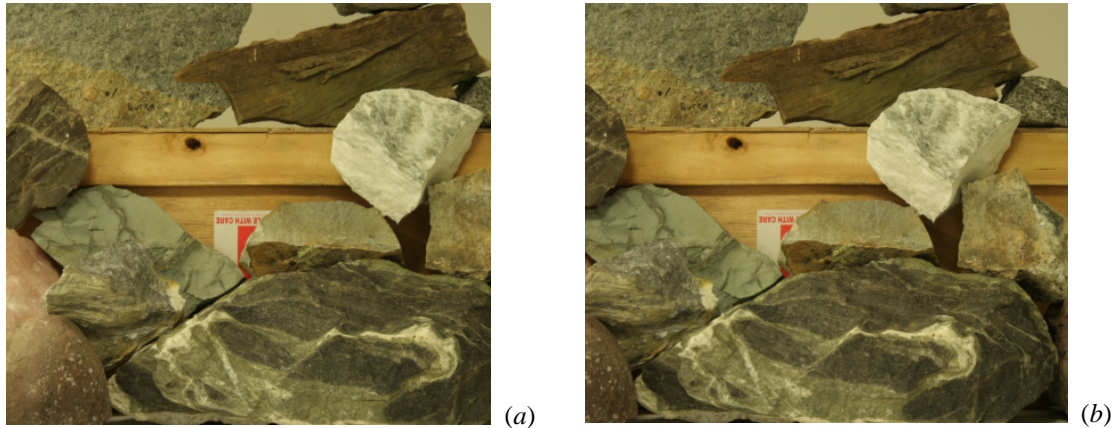


Figura 1.2 - (a) Imagen original estereoscópica izquierda; (b) Imagen original estereoscópica derecha. Como se puede observar únicamente aparece un ligero desplazamiento horizontal debido a la posición de las cámaras

1.2. Geometría del sistema estereoscópico

Un sistema convencional está caracterizado por un par de cámaras con sus ejes ópticos (Z_I y Z_D) mutuamente paralelos y separados por una distancia horizontal que se denomina *línea base*, en la Figura 1.3 se identifica con el parámetro b . Las cámaras tienen sus *ejes ópticos* perpendiculares a la línea base y sus *líneas de exploración o epipolares* paralelas a la línea base. Las líneas epipolares son líneas que unen un mismo punto en la escena en las imágenes izquierda y derecha.

Como se puede ver en la Figura 1.3 en este sistema de ejes ópticos paralelos el desplazamiento entre los centros ópticos de las dos cámaras es horizontal, esto se traduce en el hecho de que las imágenes de un punto determinado de la escena captado por ambas cámaras difiere solamente en la componente horizontal.

Las Figura 1.3 muestra la geometría de un par de cámaras en estéreo, representadas por sus modelos puntuales con sus planos imagen, I_I e I_D reflejados sobre sus centros de proyección, O_I y O_D , respectivamente.

El origen del sistema de coordenadas de referencia o del mundo está en O , siendo la *longitud focal efectiva* de cada cámara f , y la línea base b como ya se ha definido anteriormente. Los ejes de coordenadas del mundo X , Y , y Z se sitúan entre los ejes de ambas cámaras. Como consecuencia de la geometría de la imagen se obtiene la denominada *restricción epipolar*, que ayuda a limitar el espacio de búsqueda de correspondencias, de manera que en el sistema de ejes paralelos convencional todos los planos epipolares originan líneas horizontales al cortarse con los planos de las imágenes. En un sistema con la geometría anterior se obtiene un valor de *disparidad* d , para cada par de puntos emparejados $P_I(x_I, y_I)$ y $P_D(x_D, y_D)$ dado por $d = x_I - x_D$.

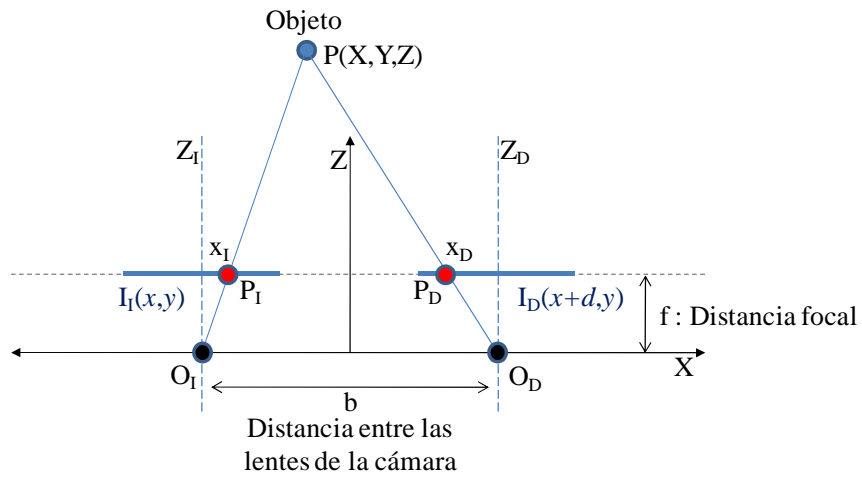


Figura 1.3 - Representación de la proyección estereó. Geometría de dos cámaras en estereó con ejes ópticos paralelos desde una perspectiva superior

1.3. Obtención de la distancia

Considerando una relación geométrica de semejanza de triángulos, las coordenadas del punto de la escena $P(X,Y,Z)$ pueden deducirse fácilmente sin más que observar la Figura 1.3, obteniendo los resultados dados por la ecuación (1.1). Se deduce a partir de la ecuación (1.1) que cuando se utiliza esta geometría, la profundidad Z , es inversamente proporcional a la disparidad de la imagen y para una profundidad dada, a mayor b mayor d .

$$\left. \begin{array}{l} O_I : \frac{\frac{b}{2} + X}{Z} = \frac{x_I}{f} \\ O_D : -\frac{\frac{b}{2} - X}{Z} = \frac{x_D}{f} \end{array} \right\} \Rightarrow \left. \begin{array}{l} x_I = \frac{f}{Z} \left(X + \frac{b}{2} \right) \\ x_D = \frac{f}{Z} \left(X - \frac{b}{2} \right) \end{array} \right\} \Rightarrow d = x_I - x_D = \frac{fb}{Z} \Rightarrow Z = \frac{fb}{d} \quad (1.1)$$

1.4. Motivación de la investigación

El trabajo que se presenta tiene su origen en las actividades de investigación planteadas dentro del grupo ISCAR (2010) destacando los siguientes proyectos actualmente en vigor:

- 1) Plataforma de Planificación, Simulación y Sistema de Vigilancia, Búsqueda y Rescate en el Mar mediante colaboración de Vehículos Autónomos Marinos y Aéreos (DPI2009-14552-C02-01), 2009-2012, perteneciente al Plan Nacional de I+D+i.
- 2) FONCYCIT (Unión Europea-México) (2009-2011) Análisis de Imágenes para el Control de Robots Autónomos, con participación de la Universidad Complutense, el Instituto Politécnico Nacional de Méjico, la Universidad de Guadalajara en México y la Universidad Libre de Berlín.

Con anterioridad, el grupo ha desarrollado actividades de investigación en el marco de otros dos proyectos en el ámbito aeroespacial con la empresa TCP Sistemas e Ingeniería titulados:

- 1) AUTO-ROVER: estudio de Autonomía basada en imágenes para Rover de exploración planetaria.
- 2) Visión Estereoscópica para AUTOROVER: Investigación aplicada de Autonomía basada en imágenes para “Rover” de Exploración planetaria.

En todos ellos la visión estereoscópica constituye uno de los componentes esenciales ya que su orientación es hacia la navegación autónoma de vehículos, tanto aéreos como terrestres no tripulados, los cuales están dotados con los correspondientes sensores ópticos de visión artificial. En todos ellos el principal problema que se plantea

es el de la correspondencia estereoscópica, máxime teniendo en cuenta que operan en entornos de exterior donde los problemas relacionados con el procesamiento de las imágenes se complican enormemente en relación a los entornos de interior.

1.5. Planteamiento del problema

En este apartado se plantea el problema así como los objetivos para el desarrollo de un método de visión estereoscópica para un sistema en tiempo real.

Hoy en día existe una gran cantidad de trabajos que intentan simular la visión humana. Cada uno de ellos se ha aplicado a distintas necesidades tales como entrenamiento médico, educación, astronomía, biología, entre otras.

Se trata por lo tanto de encontrar un método para la obtención de la disparidad, el cual proporcione buenos resultados en el menor tiempo posible. Además debe tener en cuenta el problema de las oclusiones, ya que al estar instalado en un sistema de navegación autónomo será un método que no necesite intervención del usuario. La oclusión u ocultamiento, es el fenómeno por el cual un objeto, o parte de éste, no aparece bien definido en alguna de las dos imágenes del par estéreo, mientras que en la otra imagen sí está presente. Este efecto viene provocado por la geometría del sistema estéreo, ya que al estar ambas cámaras desplazadas en el eje horizontal puede que una misma zona se vea ocultada en alguna de las imágenes.

Cualquier sistema de visión estereoscópica consta de los siguientes pasos fundamentales: 1) captura de imágenes; 2) geometría del sistema; 3) extracción de características; 4) correspondencia de características y 5) obtención de la tercera dimensión. De todas ellas, es la correspondencia estereoscópica la que entraña la mayor dificultad y donde se han concentrado los mayores esfuerzos por parte de la comunidad científica. Este trabajo se centra precisamente en él con el fin de tratar de encontrar una solución satisfactoria para su aplicación a entornos reales de exterior.

1.5.1. Objetivos generales

- 1) Aprender a manejar referencias bibliográficas, así como la forma de abordar las investigaciones.
- 2) Identificar líneas de investigación futuras.

- 3) Difundir los resultados mediante publicaciones científicas u otro tipo de divulgación.

1.5.2. Objetivos específicos

- 1) Aprender y familiarizarse con técnicas de visión estereoscópica, centrándose en los métodos relativos a la correspondencia estereoscópica.
- 2) Analizar el problema y determinar cuáles de los métodos anteriores o combinaciones de los mismos podrían aplicarse a la resolución del problema de la estima automática de la disparidad obtenida del par de imágenes estereoscópicas.
- 3) Analizar las ventajas e inconvenientes de cada uno de ellos así como su problemática.
- 4) Diseñar una estrategia de solución a un problema real.

1.6. Aportaciones a la investigación

Como resumen de todo lo anterior, las aportaciones de investigación realizadas en el presente trabajo son las siguientes:

- 1) Análisis comparativo de algoritmos para resolver el problema de la correspondencia estereoscópica y obtención de mapas de disparidad.
- 2) Análisis de métodos de filtrado del mapa de disparidad para mejorar los resultados de los algoritmos de correspondencia.
- 3) Diseño específico de una técnica de filtrado basada en la propagación de la disparidad.
- 4) Sentar las bases de trabajos futuros y proponer futuras vías de investigación en base a los resultados obtenidos para su extensión a entornos reales no estructurados.

1.7. Organización de la memoria

La presente memoria se organiza en cinco capítulos. En este capítulo se han expuesto los objetivos que se plantearon en la investigación aquí recogida, en base a los planteamientos previos y que tratan de dar solución a la problemática existente. En el capítulo dos se realiza un estudio del estado del arte en torno a la visión estereoscópica y a los métodos utilizados para realizar la correspondencia y obtener la disparidad. En el capítulo tres se describen los métodos usados así como la solución ofrecida en esta investigación y aportación principal. Será en el capítulo cuatro donde se estudien los resultados obtenidos con los nuevos algoritmos, siendo estos comparados con los métodos estudiados en la literatura. Finalmente en el capítulo cinco se extraen las conclusiones pertinentes y se exponen las líneas de investigación futuras, que dan pie al inicio de la investigación en los estudios de Doctorado.

REVISIÓN DE MÉTODOS

2.1. Visión general

Con la visión estereoscópica se obtiene una forma de inferir profundidad usando dos imágenes (par estéreo), considerando una tomada por una cámara identificada como derecha y otra mediante una cámara izquierda (Srinivasan y Malassiotis, 1999). La finalidad básica consiste en extraer el mapa de disparidad para obtener así la profundidad, teniendo en cuenta que la geometría de la cámara es previamente conocida, ecuación (1.1). Esta disparidad se define como las diferencias espaciales entre las imágenes del par estéreo y será considerada inversamente proporcional a la profundidad. El problema de correspondencia se definirá mediante la búsqueda de la correspondencia correcta de un punto, y es considerado una de las partes centrales y más difíciles del problema estéreo.

Como se ha mencionado previamente, los procesos de visión estereoscópica se estructuran en cinco grandes pasos, siendo la correspondencia el más complejo y el que ha suscitado la mayor parte en la investigación relacionada con la visión estereoscópica, siendo además el que constituye el objetivo del presente trabajo.

Por tanto, se asume que las imágenes están rectificadas (Faugeras, 1993). Dado un par de imágenes estereoscópicas y con los parámetros de orientación de las cámaras previamente conocidos (Trucco y Verri, 1998), se supone que los puntos correspondientes están en la misma línea epipolar en ambas imágenes, por lo que las cuestiones relacionadas con la captura y geometría de las mismas quedan fuera del ámbito de este trabajo, si bien constituyen el fundamento de algunos trabajos futuros como se verá posteriormente.

Centrándonos ya sobre los métodos de correspondencia, los métodos de relajación estocásticos han recibido una especial atención (Chang y Chatterjee, 1990; Gamble y col., 1989). Estas técnicas intentan obtener una solución óptima al problema de la correspondencia sin caer en un máximo local. Un ejemplo de estos métodos es el algoritmo de *Simulated Annealing*, que se usa en esta investigación, habiendo sido

aplicado previamente en Pajares y Cruz (2004). Por otra parte, la idea de analizar imágenes a varias resoluciones ha recibido mucha atención en el campo de la visión estéreo. El enfoque de multirresolución es un modo de representar los datos efectiva y eficientemente con el objetivo de reducir la complejidad computacional (Barnard, 1989; Lew y col., 1992).

La principal guía para resolver el problema de la correspondencia de un par de imágenes estéreo, se basa en la semejanza o la no semejanza, por ello, las características o áreas correspondientes deben ser similares en ambas imágenes del par estéreo. Existen diferentes medidas de semejanza, difiriendo entre sí en el coste computacional, la eficiencia, etc.

En visión estereoscópica se pueden distinguir dos grupos de técnicas para establecer la correspondencia a partir de dos imágenes estereoscópicas: *basadas en el área* (“*area-based*”) y *basadas en las características* (“*feature-based*”).

Los métodos basados en características restringen la búsqueda a un conjunto disperso de características. Se emplean propiedades simbólicas y numéricas de las características, obtenidas a partir de los llamados descriptores, los cuales se encargan de procesar y extraer características de una determinada imagen.

- 1) En los métodos basados en área, los elementos a comparar son ventanas de la imagen de dimensión fija, y el criterio de semejanza es una medida de la correspondencia entre las ventanas de las dos imágenes. El elemento correspondiente queda determinado por la ventana que maximiza el criterio de semejanza dentro de la región de búsqueda. Dentro de este segundo grupo se pueden distinguir métodos como: la suma de diferencias absolutas (*Sum of Absolute Differences: SAD*), la suma de diferencias al cuadrado (*Sum of Squared Differences: SSD*), la correlación cruzada normalizada (*Normalized Cross Correlation: NCC*) y el coeficiente de correlación de Pearson (López-Valles y col., 2005; Pajares y Cruz, 2007; Kanade y Okutomi, 1994; Okutomi y Kanade, 1993; Matthies y col., 1988, 1989; Hirschmüller y col., 2002).

En este capítulo se van a estudiar dichas técnicas bajo un estudio realizado en la literatura relacionada. Para ambos métodos se comenzará con una introducción general y se expondrán tanto las ventajas como los inconvenientes respecto a su aplicación.

2.2. Métodos basados en el área

Los métodos basados en el área utilizan información que tienen en cuenta la intensidad de la vecindad local de un píxel en una imagen, y la información de intensidad en una vecindad homóloga del que se considera el mismo píxel en la otra imagen del par estereoscópico (Cochran y Medioni, 1992; Scharstein y Szeliski, 2002). Se pueden encontrar dos tipos de enfoques: el primero hace referencia al coeficiente de correlación y el segundo a las medidas estadísticas, usadas generalmente para identificar texturas.

El problema planteado con esta técnica consiste en encontrar el píxel que se ajusta de manera más adecuada al píxel utilizado como estudio en la imagen de referencia del par estereoscópico. La epipolaridad tiene un papel muy importante, ya que reduce esta búsqueda a un problema de búsqueda en una dimensión.

La correspondencia se establece por similitud entre las propiedades de los píxeles bajo el análisis realizado dentro de la vecindad local. Pero hay que tener en cuenta que dependiendo del criterio utilizado los resultados pueden ser dispares. Por ejemplo, la suma de diferencias de cuadrados (*SSD*, *Sum of Squared Differences*) de la ecuación (2.1) es muy sensible a diferencias de iluminación, por lo que en su lugar se utiliza la suma de diferencias de cuadrados centradas respecto a la media (Pajares y Cruz, 2007) (2.2). El SSD ha sido utilizado en diferentes trabajos sobre visión estereoscópica basados en secuencias de imágenes (Kanade y Okutomi, 1994; Okutomi y Kanade, 1993; Matthies y col., 1988, 1989), donde la medida SSD es considerada una variable aleatoria.

$$\sum_{i=1}^M \sum_{j=1}^N (I_I(i, j) - I_D(i + d, j))^2 \quad (2.1)$$

$$\sum_{i=1}^M \sum_{j=1}^N \left(\frac{I_I(i, j) - \mu_I}{\sigma_I} - \frac{I_D(i + d, j) - \mu_D}{\sigma_D} \right)^2 \quad (2.2)$$

El método estadístico estándar más utilizado es el coeficiente de correlación de Pearson (2.3). Este coeficiente mide la relación lineal entre dos variables cuantitativas, y a diferencia de la covarianza, es independiente de la escala de medición de las variables.

$$C = \frac{\sigma_{ID}^2}{\sqrt{\sigma_I^2 \sigma_D^2}} \quad (2.3)$$

Para que la correspondencia sea robusta, es necesario considerar una ventana de vecindad, ya que si sólo se tiene en cuenta la intensidad del píxel a evaluar se considerará poco fiable al poder aparecer varios píxeles con la misma intensidad sobre la línea epipolar (Pajares y Cruz, 2007). Existen algunos problemas relacionados con el tamaño de la ventana:

- 1) Si la ventana tiene una dimensión grande se reducen los problemas derivados del ruido pero aumenta el coste computacional. El trabajar con una ventana de dicha dimensión puede violar el principio de continuidad de la disparidad, en el sentido de que los píxeles pertenecientes a la ventana de vecindad pueden presentar niveles de disparidad y por tanto de intensidad muy diferentes, debido a la presencia de oclusiones o discontinuidades. Por ello, una ventana demasiado grande puede originar falsas correspondencias y pérdidas de resolución cuando queremos aplicarlo a una reconstrucción 3-D.
- 2) El trabajar con una ventana cuya resolución sea pequeña puede no ser suficiente para capturar la estructura de la imagen, presentando a su vez una sensibilidad alta al ruido, lo cual nos puede llegar a originar falsas correspondencias.

Shirai (1987) propone un algoritmo que comienza con una ventana de dimensiones pequeñas y si se produce claramente un mínimo, de la función que se esté utilizando para la evaluación, por debajo de un umbral fijado, la correspondencia se acepta. Si es mayor que un segundo umbral prefijado la correspondencia se rechaza. Y si la decisión no es clara se realiza el proceso con una ventana de mayores dimensiones.

En el trabajo realizado por Levine y col. (1973) se propone la utilización de una ventana adaptativa cuya dimensión a lo largo de la línea epipolar para un punto determinado varía inversamente con la intensidad en ese punto. El objetivo es elegir la ventana de menor dimensión y que a la vez sea apropiada para la correspondencia. A su vez, en los estudios de Kanade y Okutomi (1994) se puede destacar una técnica con una

ventana adaptativa en la que la dimensión y la forma de la ventana se seleccionan estudiando la variación de la intensidad y la disparidad (estimada con anterioridad).

Ventajas de las técnicas basadas en el área:

- 2) La correspondencia se lleva a cabo píxel a píxel siguiendo las líneas epipolares. Esto no requiere de conocimiento previo sobre si el píxel pertenece a un determinado objeto o no.
- 3) Obtiene buenos resultados sobre imágenes con texturas (López-Valles y col., 2005).
- 4) Permiten crear mapas densos de disparidad, es decir, se obtendrá una disparidad para todos los puntos de la escena, lo cual resulta útil para reconstrucciones de superficies (López-Valles y col., 2005).
- 5) Son fáciles de paralelizar utilizando lenguajes de descripción sobre hardware configurable o bien utilizando tarjetas aceleradoras gráficas GPU las cuales permiten operaciones con matrices de gran tamaño (Moravánszky, 2004; Fatahalian y col., 2004).

Desventajas de las técnicas basadas en el área (López-Valles y col., 2005):

- 1) De la experiencia se puede deducir cómo los objetos que son homogéneos hacen que la correspondencia píxel a píxel se convierta en una difícil tarea a la hora de decidir cuál es el píxel homólogo en la otra imagen. Esto se debe a que varios píxeles pueden tener la misma información de vecindad y por consiguiente hay que definir alguna regla para saber cuál de ellos es el correcto.
- 2) Presentan problemas con imágenes con elevadas discontinuidades de superficie.
- 3) Son muy sensibles a variaciones fotométricas debidas a posibles sombras u otros fenómenos.
- 4) Requieren un proceso posterior para la eliminación de falsas correspondencias.
- 5) Plantean graves problemas con las oclusiones.

2.3. Métodos basados en características

Como contraposición al hecho de establecer correspondencias punto a punto, surge el uso de características, que atesoran un conjunto de propiedades invariantes a la proyección en mayor o menor medida. Se entiende por característica alguna estructura significativa de la imagen, en concreto: *puntos de borde*, *segmentos de borde* (rectos o curvilíneos) o *regiones*. Los métodos basados en las características (Tang y col., 2002) calculan atributos para esas características (píxeles o regiones) bajo correspondencia; pueden ser simples atributos, tales como el color de los píxeles (Klaus y col., 2006) o propiedades obtenidas aplicando algún operador como el módulo del gradiente (Klaus y col., 2006) y el ángulo del gradiente o la Laplaciana (Lew y col., 1994).

Los métodos basados en las características utilizan normalmente conjuntos de píxeles con atributos similares, ya sean píxeles pertenecientes a bordes (Tang y col., 2002; Grimson, 1985; Ruichek y Postaire, 1996), segmentos de borde (Medioni y Nevatia, 1985; Pajares y Cruz, 2004, 2006; Scaramuzza y col., 2008), regiones (McKinnon y Baltes, 2004; Marapane y Trivedi, 1989; Baykant, 2008), esquinas (Wang y col., 1983), o enfoques jerárquicos (Wei y Quan, 2004) donde primero se establece la correspondencia entre bordes o esquinas y después las regiones. En Baykant (2008) se plantean dos tipos de enfoques: en el primero se utiliza la minimización del error global de energía mediante funciones de suavizado, y en el segundo, utiliza el crecimiento de regiones para tener el menor error de energía en un determinado umbral. En Tang y col. (2002) se utilizan regiones para realizar la correspondencia considerando los siguientes atributos específicos: área, centroide y ángulos. En el estudio realizado por Venkateswar y Chellappa (1995) se propone un algoritmo de correspondencias basado en características explotando cuatro tipos de éstas: líneas, vértices, contornos y superficies, donde la correspondencia comienza en el nivel más alto de la jerarquía, las superficies, y continúa hasta el más bajo, o sea, las líneas. Una vez que los componentes de la jerarquía más alta han sido puestos en correspondencia, las características que lo componen no se vuelven a analizar en los niveles más bajos. Otra aproximación basada en características consiste en segmentar las imágenes primero en regiones y después emparejar dichas regiones (Wei y Quan, 2004).

Muchos de estos métodos tienen en cuenta los píxeles y su vecindad, por tanto, desde este punto de vista podrían considerarse basados en el área. Como ya se ha

definido anteriormente, el color es otro de los atributos que puede utilizarse a nivel de píxel individual o considerar esta característica como una intervención de un entorno de vecindad (Klaus y col., 2006) por lo que este tipo de técnicas podrían considerarse híbridas. En Klaus y col. (2006) se utiliza la segmentación del color en la imagen de referencia, realizando una coincidencia auto-adaptativa del resultado, lo cual va a maximizar el número de correspondencias fiables y en lugar de asignar un valor de disparidad a cada píxel de la imagen, se le asignará un plano de disparidad a cada segmento.

Otra forma de abordar el problema de la correspondencia de características entre imágenes es mediante la minimización de una función de coste. La programación dinámica es una forma eficiente de minimizar (o maximizar) funciones de gran número de variables discretas, ya que plantea el problema de correspondencia estereoscópica como el trazado de un camino sobre un espacio de búsqueda bidimensional. En Baker y Binford (1981) y Ohta y Kanade (1985) se han realizado intentos satisfactorios utilizando programación dinámica para resolver el problema de la correspondencia estéreo. En ambos casos se utilizan los contornos como características básicas.

Ventajas de las técnicas basadas en las características:

- 1) Las características ofrecen información más fiable y completa que los niveles de intensidad.
- 2) Permiten utilizar restricciones geométricas entre las características.
- 3) Son muy robustas y rápidas.
- 4) Son relativamente insensibles a cambios de iluminación.

Desventajas de las técnicas basadas en las características (López-Valles y col., 2005):

- 1) Proporcionan información dispersa, ya que limitan las regiones a analizar a características fiables de las imágenes, como pueden ser los contornos, las curvas, etc.
- 2) Ofrece mapas de disparidad escasos, ya que reduce la densidad de puntos sobre los que se estima la profundidad (útiles para aplicaciones de navegación).

En la última década los métodos jerárquicos y la segmentación basada en regiones ha recibido especial atención (López-Valles y col., 2005).

En Pajares y Cruz (2007) se menciona que los métodos basados en las características, en general, primero establecen correspondencias *locales* en base a la restricción de similitud y luego verifican la solidez de las mismas mediante ciertas estrategias de naturaleza *global*. Esta técnica se conoce como proceso de *relajación*, la cual no solamente tiene en cuenta el valor de la correlación obtenida para los píxeles de la línea que se analiza, sino que también otorga importancia a los valores de correlación obtenidos para una cierta vecindad. Algunos de los procedimientos para llevar a cabo este objetivo son: *a)* basados en procesos de *optimización*, *b)* de naturaleza *probabilista* o *c)* basados en redes de naturaleza *causal*.

De manera general, una vez obtenido el mapa de disparidad, el objetivo se centra ahora en elegir un mecanismo apropiado para filtrar dicho mapa mediante la aplicación de la restricción de la continuidad de la disparidad (Pajares y Cruz, 2004), es decir, los píxeles que pertenecen a un mismo objeto deben poseer niveles de disparidad similares. Tanto los procesos que aplican un filtrado de naturaleza estadística como los que ajustan planos a los valores de disparidad obtenidos, son procesos globales que consideran el conjunto de píxeles en la región de donde extraen los valores. Lo más conveniente consiste en estudiar las contribuciones de los píxeles vecinos, es decir, cada píxel vecino contribuirá de modo diferente a la modificación de los valores de disparidad en función de la relación con el píxel que se está tratando. En esta línea se encuentran los estudios realizados bajo el paradigma de los denominados Mapas Cognitivos Fuzzy o “*Fuzzy Cognitive Maps*” (*FCMs*) (Pajares y Cruz, 2006), la Red Neuronal de Hopfield o “*Hopfield Neural Network*” (*HNN*) (Pajares y col., 1998) y el Enfriamiento Simulado o “*Simulated Annealing*” (*SA*) (Pajares y col., 2004).

PROPUESTA DE MÉTODOS

3.1. Introducción y medidas del error

En este capítulo se abordan distintos métodos y técnicas de procesamiento de imágenes estereoscópicas que se han utilizado como base en el desarrollo de la investigación. En concreto, se estudian los métodos basados en la correlación de Pearson, así como la minimización del error global de energía y la correspondencia estéreo basada en línea de crecimiento propuestas en Baykant (2008), y la correspondencia estéreo basada en la segmentación, estudiada en Klaus y col. (2006) e implementada por Lankton (2010). Además, se estudian distintos tipos de filtrado que pueden ser aplicados a las imágenes obtenidas mediante las técnicas anteriores.

Para poder comparar los algoritmos que se van a estudiar a continuación, se ha utilizado el porcentaje de error. Se trata de una simple comparación entre la disparidad obtenida mediante estas técnicas y la imagen de disparidad que se ha considerado como válida, en base a la información proporcionada por los suministradores de las mismas, como se verá posteriormente. Con ello, se obtiene el porcentaje de error que se comete al aplicar cada técnica, contando el número de píxeles divididos entre el número de píxeles total (ecuación 3.1)

$$\sum_{i=1}^M \sum_{j=1}^N \frac{e}{M \cdot N} \quad (3.1)$$

donde e es el número de errores que cumple: $real(i, j) \neq obtenido(i, j)$

Además de dicha medida de comparación, se utiliza el Error Cuadrático Medio. Esta función proporciona la medida de las diferencias en promedio de los valores obtenidos y los observados a partir de las imágenes de disparidad real (ecuación 3.2). De esta forma no sólo se puede comparar si son exactamente iguales ambas imágenes, sino que además da un valor del error medio al cuadrado cometido por cada píxel de la imagen. Esto hace que aunque el porcentaje de error sea mayor en una determinada técnica, si su error cuadrático medio es inferior, dicha técnica pueda ser mejor que aquella que obtiene un porcentaje de error menor y mayor error cuadrático medio.

$$\sum_{i=1}^M \sum_{j=1}^N \frac{(real(i, j) - obtenido(i, j))^2}{M \cdot N} \quad (3.2)$$

Por ello, hay que buscar el equilibrio correcto entre el porcentaje de errores y el error cuadrático medio, intentado que ambos sean lo más pequeño posible.

3.2. Correlación de Pearson

Se trata de un método basado en el área, por lo tanto se trabajará con una ventana de vecindad $M \times N$ en torno al punto para el cual se está buscando la correspondencia. Antes de comenzar los cálculos es necesario elegir una de las dos imágenes del par estéreo, ésta será utilizada como referencia para buscar en la otra imagen el punto cuyo valor de correlación sea el máximo dentro de un rango $dmax$ de puntos de la línea epipolar.

En la Figura 3.1 se puede observar cómo un punto (x_d, y_d) de la imagen derecha, aparece en la imagen izquierda desplazado hacia la derecha con respecto su posición en la imagen de referencia, es decir, un punto de la imagen derecha (x_d, y_d) se corresponde con otro punto de la imagen de la izquierda (x_i, y_i) con un cierto desplazamiento d sobre el eje x de la línea epipolar. Por lo tanto, un punto (x_d, y_d) de la imagen de la derecha se corresponde con un punto $(x_d + d, y_d)$ en la imagen de la izquierda, siendo este d el valor de disparidad que se desea obtener, y que se corresponde en este caso particular con el valor máximo de correlación dentro del rango $dmax$.

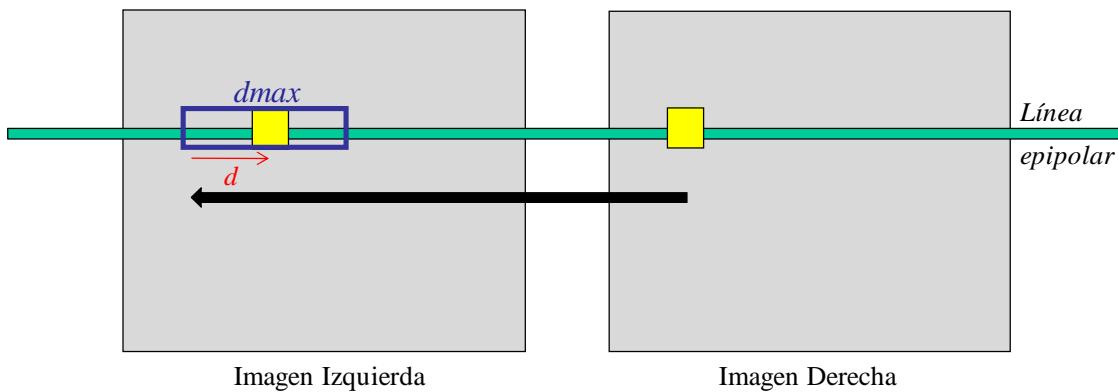


Figura 3.1 - Rango $dmax$ utilizando como referencia la imagen derecha

En la Figura 3.2 se muestra un diagrama de flujo sobre los pasos a seguir en la aplicación del método de correlación.

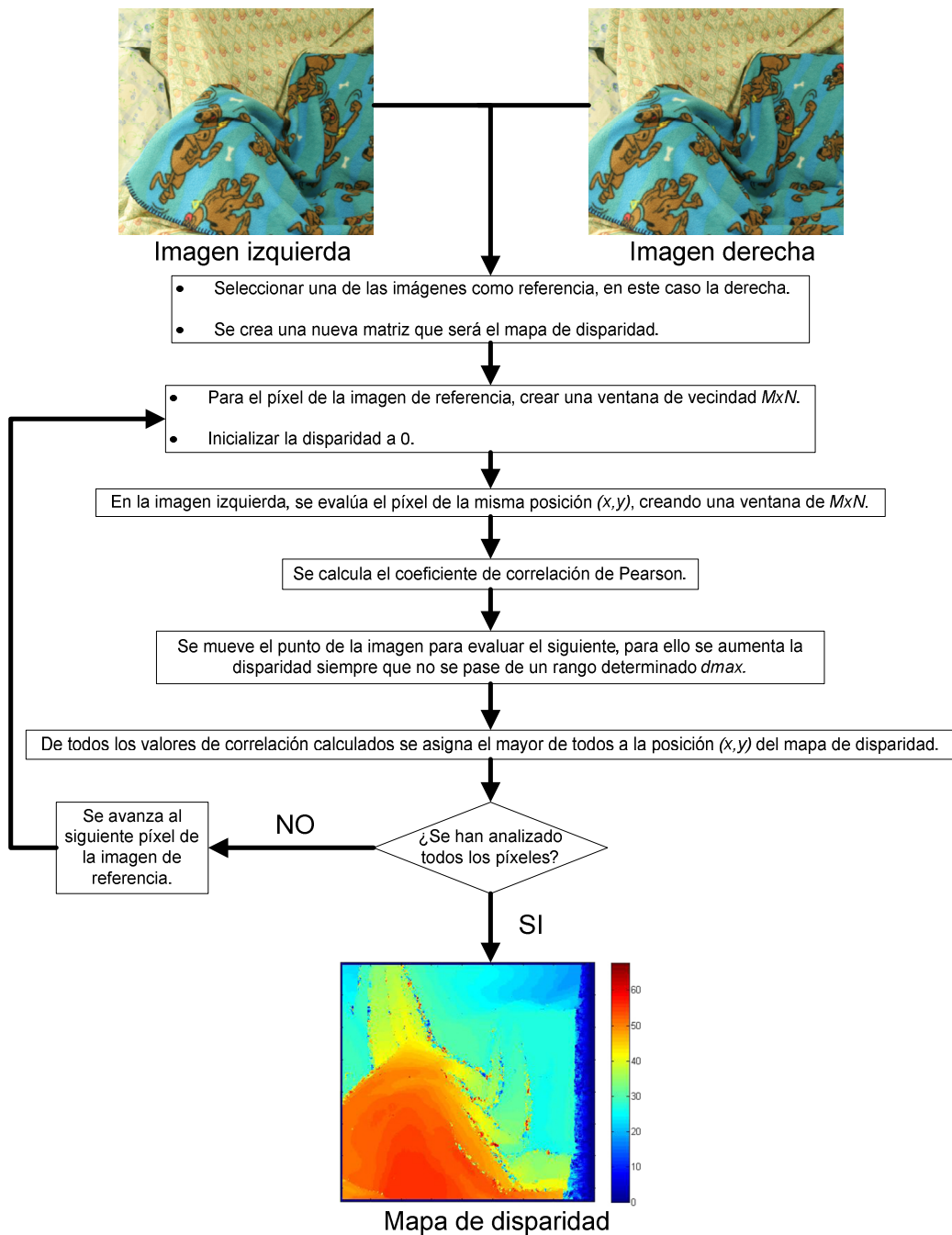


Figura 3.2 - Esquema de los pasos a seguir en el algoritmo de correlación.

De este modo, tomando como referencia la imagen derecha y siguiendo los pasos de la Figura 3.2, el proceso evoluciona como sigue:

1. Para el punto (píxel) a evaluar de la imagen derecha, se crea una ventana de vecindad con dimensiones $M \times N$. Se inicializa el valor d del rango en 0, $d = 0$.
2. En la imagen izquierda se evalúa la posición $(x_d + d, y_d)$ por lo que se abre una ventana de las mismas dimensiones que el punto de la derecha ($M \times N$).

3. Se calcula el coeficiente de correlación de Pearson según la ecuación (2.3) donde los subíndices I y D se refieren a las imágenes izquierda y derecha, respectivamente, σ_I^2 y σ_D^2 (3.3) representan la varianza de los niveles de intensidad en las correspondientes ventanas de vecindad y σ_{ID}^2 (3.4) es la covarianza de los niveles de intensidad entre las ventanas izquierda y derecha. La intensidad de cada píxel (i,j) viene dada por $I_k(i,j)$ y la media en el entorno de vecindad o ventana resulta ser μ_k . En ambos casos, el subíndice k indica si se trata de la imagen izquierda o derecha.

$$\sigma_k^2 = \sum_{i=1}^M \sum_{j=1}^N \frac{(I_k(i,j) - \mu_k)^2}{M \cdot N}; \quad k = I, D \quad (3.3)$$

$$\sigma_{ID}^2 = \sum_{i=1}^M \sum_{j=1}^N \frac{(I_I(i,j) - \mu_I)(I_D(i,j) - \mu_D)}{M \cdot N} \quad (3.4)$$

4. Se almacena el valor de correlación obtenido y se avanza una posición el punto (x_i, y_i) de la imagen izquierda, es decir, se aumenta el valor d ($d=d+1$) siempre que d sea menor que $dmax$, para poder evaluar el siguiente punto (x_{d+d}, y_d) dentro del rango $dmax$, y se realiza de nuevo el paso 2.
5. Una vez se han calculado todos los valores del rango $dmax$, se obtiene la d que ha obtenido el mayor valor de correlación, lo que indicará el valor de disparidad para el cual existe una mayor dependencia de variables, es decir, indica cuál de todos los puntos del intervalo de la imagen izquierda es el que más se parece al que se está evaluando en la imagen derecha.
6. Se asigna este valor máximo, al punto (x_d, y_d) de una nueva matriz que será el *mapa de disparidad*. Esta matriz será del mismo tamaño que la imagen derecha en este caso, que es la que se utiliza como referencia, aunque en general ambas imágenes tendrán el mismo tamaño.
7. Si no se han analizado todos los puntos de la imagen derecha se vuelve al paso 1 para seguir evaluando el siguiente punto. En el caso de que todos los puntos hayan sido evaluados, ya se tendría el mapa de disparidad correspondiente a las dos imágenes del par estéreo.

Hay que tener en cuenta que se debe dejar un marco (borde) alrededor de las imágenes, esto es debido a la ventana de vecindad. Dado que cada ventana tiene unas dimensiones $M \times N$ fijas, para que todos los puntos tengan el mismo número de vecinos, se debe dejar este borde, lo que hace que para los extremos de las imágenes no se obtenga ninguna disparidad.

Además, cabe destacar, que este algoritmo hay que aplicarlo a cada uno de los canales de color que posee la imagen. Por lo que si se trabaja con imágenes en blanco y negro únicamente se tiene un canal, pero si se trabaja con imágenes a color, hay que aplicarlos a los canales rojo, verde y azul (R,G,B). Esto no supone un problema, simplemente en la evaluación de la misma ventana de vecindad, hay que calcular los valores para los 3 canales, quedando las ecuaciones (3.3) y (3.4) como las ecuaciones (3.5) y (3.6) respectivamente:

$$\sigma_k^2 = \sum_{i=1}^M \sum_{j=1}^N \frac{(I_k(i, j, c) - \mu_k)^2}{M \cdot N}; \quad k = I, D \quad c = 1, 2, 3 \quad (3.5)$$

$$\sigma_{ID}^2 = \sum_{i=1}^M \sum_{j=1}^N \frac{(I_I(i, j, c) - \mu_I)(I_D(i, j, c) - \mu_D)}{M \cdot N}; \quad c = 1, 2, 3 \quad (3.6)$$

Donde c representa la componente RGB de la imagen (el número de canal para el que se evalúa el punto). De esta forma, como en el coeficiente de correlación de Pearson (ecuación 2.3) hay que utilizar un único valor para la varianza y la covarianza, se puede optar por utilizar el máximo, el mínimo, o como en este caso, la media de los resultados obtenidos de los tres canales. Por lo que generalizando las ecuaciones (3.5) y (3.6) se obtienen las ecuaciones (3.7) y (3.8).

$$\sigma_k^2 = \sum_{c=1}^L \frac{\sum_{i=1}^M \sum_{j=1}^N \frac{(I_k(i, j, c) - \mu_k)^2}{M \cdot N}}{L}; \quad k = I, D \quad (3.7)$$

$$\sigma_{ID}^2 = \sum_{c=1}^L \frac{\sum_{i=1}^M \sum_{j=1}^N \frac{(I_I(i, j, c) - \mu_I)(I_D(i, j, c) - \mu_D)}{M \cdot N}}{L} \quad (3.8)$$

Donde L es el número de canales que posee la imagen. Por ello, en el caso de trabajar con imágenes a color, las ecuaciones (3.3) y (3.4) del paso 3 serán sustituidas por las ecuaciones (3.7) y (3.8)

3.3. Minimización del error global de energía

Este algoritmo (*Global error energy minimization*) es utilizado en el trabajo desarrollado por Baykant (2008). Se utiliza una técnica de correspondencia por bloques con el fin de construir una matriz con el Error de Energía por cada valor de disparidad, donde $L(i,j,c)$ y $R(i,j,c)$ representan las imágenes izquierda (Left) y derecha (Right) en formato RGB respectivamente; c representa la componente RGB de la imagen, por lo que tomará valores de $\{1,2,3\}$ para rojo (Red), verde (Green) y azul (Blue); n y m se corresponden al tamaño ($N \times M$) de la ventana del bloque de correspondencia; y d es la disparidad.

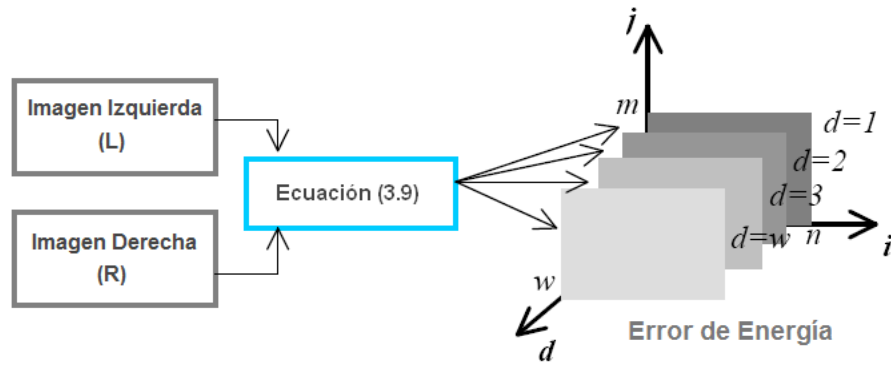
- 1) Para toda disparidad d dentro del rango de disparidad w , se calcula la matriz del error de energía, donde $e(i, j, d)$ se corresponde con el error para el píxel (i, j) con disparidad d , la cual se define mediante la ecuación (3.9) (Figura 3.3a).

$$e(i, j, d) = \frac{1}{3 \cdot N \cdot M} \cdot \sum_{x=i}^{i+n} \sum_{y=j}^{j+m} \sum_{k=1}^3 (L(x, y + d, k) - R(x, y, k))^2 \quad (3.9)$$

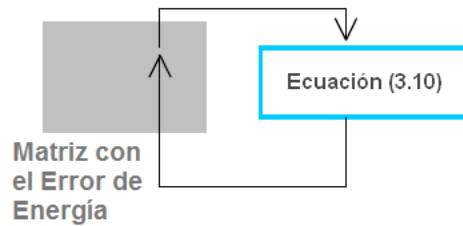
- 2) Aplicando el filtro de la media definido por la ecuación (3.10) se suaviza la imagen iterando toda la matriz del error calculada para un valor de disparidad dentro del rango de búsqueda de disparidad (Figura 3.3b). Con este procedimiento se consigue eliminar los cambios bruscos de energía, los cuales posiblemente pertenezcan a una correspondencia errónea.

$$\tilde{e}(i, j, d) = \frac{1}{N \cdot M} \cdot \sum_{x=i}^{i+n} \sum_{y=j}^{j+m} e(x, y, d) \quad (3.10)$$

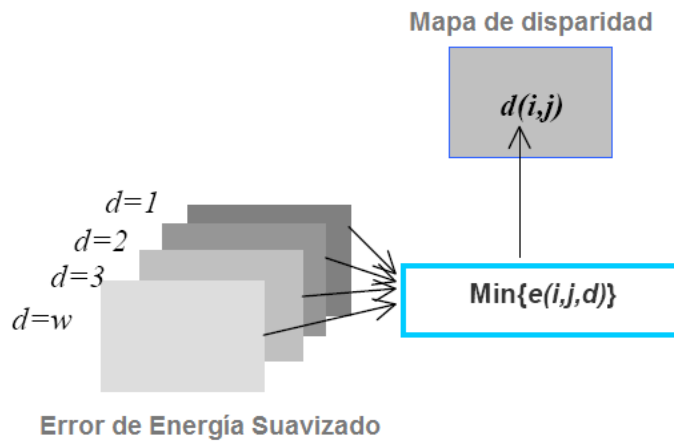
- 3) Para cada píxel (i, j) se busca la matriz con el menor error de energía $\tilde{e}(i, j, d)$ y se asigna el valor de disparidad (d) de la matriz al mapa de disparidad $d(i, j)$ generado como resultado (Figura 3.3c).



a) Construcción de la matriz de error de energía



b) Suavizado de la matriz de energía para todos los valores de disparidad



c) Generación del mapa de disparidad con el mínimo de todos los puntos

Figura 3.3 - Esquema del método usando la minimización del error global de energía mediante funciones de suavizado.

3.4. Correspondencia estereoscópica basada en línea de crecimiento

Este algoritmo (*Line growing based stereo matching*) se encuentra también propuesto en Baykant (2008). Se utiliza la técnica de crecimiento de regiones, donde en un primer paso (proceso de Selección de la Raíz o “Root Selection process”) se obtiene un punto raíz y en la segunda fase (proceso de Crecimiento de la Región o “Region Growing process”), se aumenta la región, a partir de ese punto, en función de una determinada regla.

La regla que se utiliza para la asignación de un punto raíz en la región de crecimiento consiste en seleccionar el punto que tenga un valor de energía por debajo de un determinado umbral. Para simplificar el algoritmo, se utiliza únicamente un crecimiento por filas, por lo que sólo existe un vecino para cada píxel, el cual será inspeccionado para la región de crecimiento después del punto buscado (Figura 3.4).

Los pasos a seguir son los siguientes:

1. *Proceso de Selección de la Raíz o “Root Selection process”*: se busca en las filas con el fin de encontrar un punto raíz, el cual no pertenezca a ninguna línea de crecimiento y calcular su disparidad utilizando la función de energía (ecuación 3.9) con una ventana del tipo *línea* (solamente a lo largo de la fila, $1 \times M$). Si el error de energía del punto seleccionado es menor o igual a un umbral *LineGrowingThreshold* (V_{LG}), se selecciona este punto como raíz y se avanza al paso 2. Si no, marcar ese punto como perdido (*idle*) y realizar de nuevo el paso 1 para el siguiente punto de la fila.
2. *Proceso de Crecimiento de la Región o “Region Growing process”*: calcular el error de energía del siguiente punto vecino sólo para la disparidad del punto raíz, que será llamada *región de disparidad*. Si es menor o igual que un umbral predeterminado, V_{LG} , asociar este punto a la región. De lo contrario, volver al paso 1 y encontrar un nuevo punto raíz.
3. Realizar el paso 1 y 2 fila por fila hasta completar la imagen. El crecimiento de las regiones de disparidad componen el mapa de disparidad $d(i,j)$.

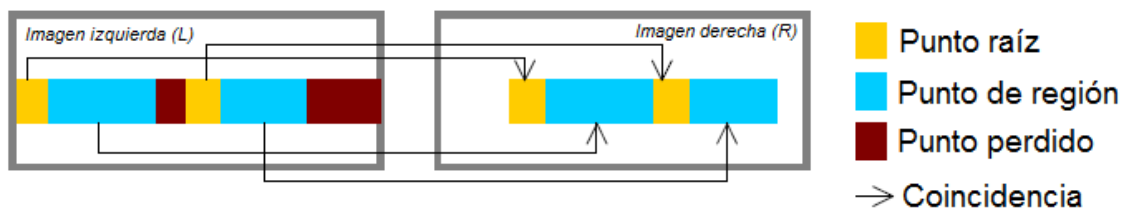


Figura 3.4 - Método usando línea de crecimiento.

3.5. Correspondencia estereoscópica basada en la segmentación

En Klaus y col. (2006) se propone un algoritmo (*Segment-based stereo matching*) que utiliza la segmentación del color en la imagen de referencia y una coincidencia auto-adaptativa del resultado que maximiza el número de correspondencias fiables, asignando un plano de disparidad a cada segmento en lugar de un valor de disparidad a cada píxel.

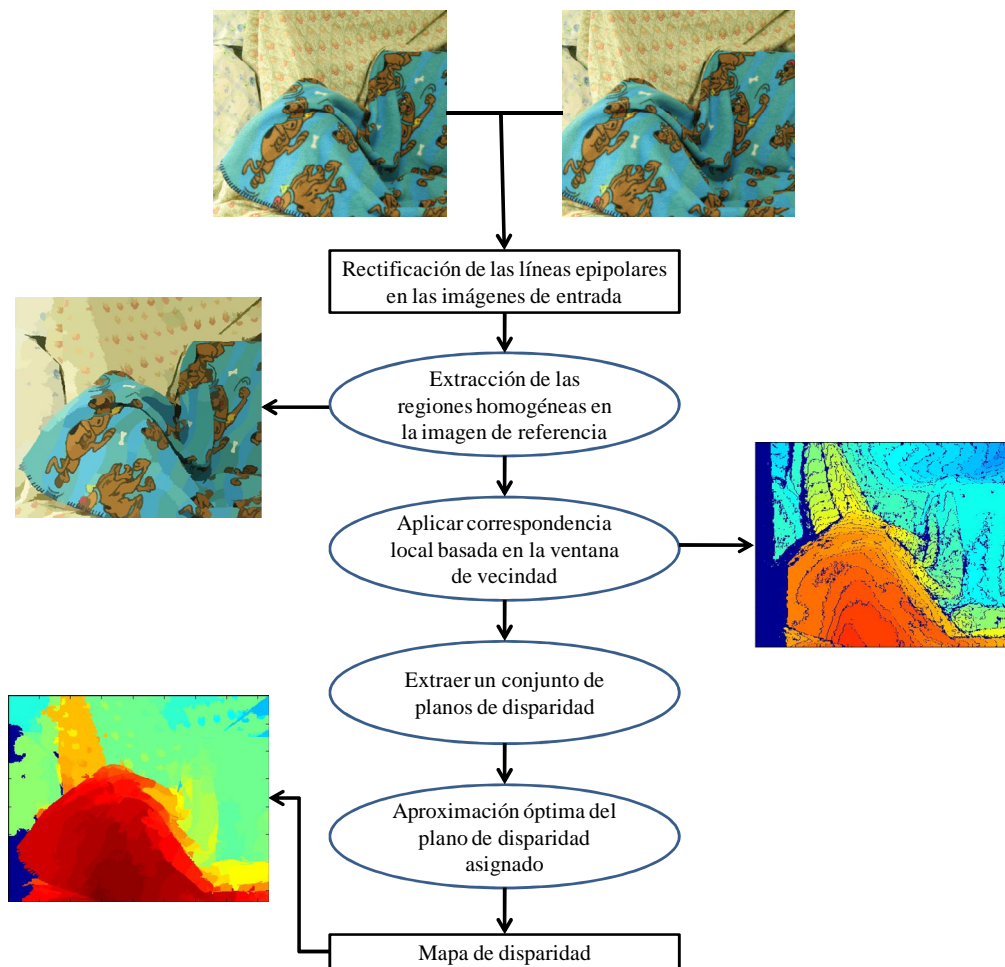


Figura 3.5 - Esquema del algoritmo de correspondencia estereoscópica basada en segmentos, con imágenes de entrada, intermedias y resultados finales del método propuesto.

El primer paso consiste en descomponer la imagen de referencia en regiones de color homogéneo o escala de grises, asumiendo que los valores en esas regiones varían suavemente y que las discontinuidades de profundidad sólo ocurren en los límites de cada región.

Este método está modelado por un conjunto de planos de disparidad. Un plano está definido por tres parámetros c_1, c_2, c_3 que determinan una disparidad d para cada píxel de la imagen de referencia $d = c_1x + c_2y + c_3$.

Debido a la gran cantidad de planos de disparidad que pueden existir, el número de éstos se reduce a la obtención de un conjunto representativo. Este procedimiento se realiza aplicando correspondencia local en el dominio del píxel seguido de una estimación del plano de disparidad.

Para la correspondencia local se define una ventana de agregación y un resultado de correspondencia. La mayoría de las medidas que tienen en cuenta la diferencia de similitud (disparidad) son: diferencias de intensidad al cuadrado (*Square Difference*) y diferencias absolutas de intensidad (*Absolute Difference*), las cuales asumen que los colores son constantes.

Otras técnicas basadas en el gradiente y las no paramétricas son más robustas frente a cambios en la ganancia de la cámara (fenómeno por el cual el color obtenido por la cámara no es el esperado), y obtienen mejores resultados. En este enfoque se utiliza una medida auto-adaptativa para la disparidad que combina la suma de diferencias absolutas de intensidad (SAD) y una medida basada en el gradiente (GRAD) (ecuaciones 3.11 y 3.12).

$$C_{SAD}(x, y, d) = \sum_{(i,j) \in N(x,y)} I_1(i, j) - I_2(i + d, j) \quad (3.11)$$

$$C_{GRAD}(x, y, d) = \sum_{(i,j) \in N_x(x,y)} |\nabla_x I_1(i, j) - \nabla_x I_2(i + d, j)| + \sum_{(i,j) \in N_y(x,y)} |\nabla_y I_1(i, j) - \nabla_y I_2(i + d, j)| \quad (3.12)$$

Donde $N(x, y)$ es una ventana 3 x 3 abierta alrededor de la posición (x, y) ; $N_x(x, y)$ es una ventana alrededor de la posición (x, y) sin la columna de más a la derecha; $N_y(x, y)$ es una ventana alrededor de la posición (x, y) sin la fila inferior; ∇_x es el gradiente en x hacia la derecha; ∇_y es el gradiente en y hacia abajo. Las imágenes en color tienen en cuenta la suma de la medida de disparidad para todos los canales (R,G,B).

Opcionalmente se introduce un peso ω entre C_{SAD} y C_{GRAD} que determina el número máximo de correspondencias fiables que son filtradas mediante una prueba de verificación cruzada (comparando los mapas de disparidad de izquierda a derecha y de derecha a izquierda) en relación con la mejor de todas las optimizaciones obtenidas (*winner-take-all*, eligiendo la disparidad que tiene menor coste de correspondencia). El resultado de la diferencia de similitudes viene dado por la ecuación (3.13).

$$C(x, y, d) = (1 - \omega) * C_{SAD}(x, y, d) + \omega * C_{GRAD}(x, y, d) \quad (3.13)$$

Además, se utiliza la fiabilidad de la correspondencia para predecir la relación señal/ruido (SNR, Signal to Noise Ratio) que es utilizada para normalizar la medida utilizada para la disparidad. Debido a esta normalización, se puede establecer un umbral de truncamiento por encima del nivel de ruido, con el fin de obtener un valor de correspondencia robusto.

Las correspondencias fiables se usan para obtener un conjunto de planos de disparidad que sean adecuados para representar la estructura de la imagen. Esto se consigue aplicando un método de ajuste robusto y una serie de pasos para refinarlo y así aumentar la precisión del plano de disparidad.

A pesar de que sólo las disparidades fiables de cada segmento son utilizadas para derivar el plano de disparidad correspondiente, el plano de estimación puede verse distorsionado por los valores extremos restantes. Una forma de determinar los parámetros del plano de disparidad es resolver un sistema de mínimos cuadrados, pero tiene el problema de que son muy sensibles a valores atípicos, de forma que soluciones lineales o la mediana son más robustas.

Este método determina una solución robusta aplicando un método de descomposición para resolver cada parámetro por separado:

- 1) El sesgo horizontal se calcula utilizando el conjunto de todas las combinaciones posibles de disparidades fiables que están en la misma línea horizontal dentro del segmento. Las derivadas $\partial d / \partial x$ se insertan en una lista y se determina una estimación robusta de la inclinación horizontal ordenando la lista y aplicando circunvolución con un núcleo Gaussiano.

- 2) El sesgo vertical se estima de manera similar, considerando todas las combinaciones posibles en la misma línea vertical.
- 3) El sesgo se utiliza para estimar un valor de disparidad robusto en el centro del segmento. Por lo tanto, las disparidades centrales de cada punto fiable son calculadas teniendo en cuenta el sesgo estimado, se insertan en una lista y se obtiene una estimación robusta.

Para aumentar la precisión del plano de disparidad se establece, por repetición, el plano adecuado para regiones agrupadas que están en el mismo plano de disparidad:

- 1) Se calcula el coste de correspondencia para cada asignación del segmento al plano. Esto se calcula sumando el coste de correspondencia para cada píxel dentro del segmento S , ecuación (3.14).

$$C_{SEG}(S, P) = \sum_{(x,y) \in S} C(x, y, d) \quad (3.14)$$

donde P es un plano de disparidad que define la disparidad d .

- 2) Se asigna a cada segmento el plano de disparidad con menor coste de correspondencia.
- 3) Se agrupan los segmentos que son asignados al plano de disparidad.
- 4) Se repite la estimación del plano para todos los grupos de segmentos.

En el paso final, se busca la solución óptima para la asignación del plano de disparidad al segmento. Por lo tanto, la correspondencia estéreo se formula como un problema de minimización de energía para el etiquetado f , que asigna a cada segmento $s \in R$ el plano correspondiente $f(s) \in D$. La energía para un etiquetado f viene dado por la ecuación (3.15)

$$E(f) = E_{data}(f) + E_{smooth}(f) \quad (3.15)$$

donde

$$E_{data}(f) = \sum_{s \in R} C_{SEG}(s, f(s)) \quad (3.16)$$

$$E_{smooth}(f) = \sum_{(\forall (s_i, s_j) \in S_N | f(s_i) \neq f(s_j))} \lambda_{disc}(s_i, s_j) \quad (3.17)$$

S_N representa el conjunto de todos los segmentos adyacentes y $\lambda_{disc}(s_i, s_j)$ es un factor de penalización que incorpora la longitud de los bordes comunes y la media de similitud del color como se propone en Bleyer y col. (2005).

3.6. Filtrado de imágenes

Como se quiere un proceso automático, y asumiendo que los algoritmos de visión estereoscópica pueden generar errores en el cálculo de la disparidad, es conveniente recurrir a filtrados de los mapas de disparidad. Estos filtrados, pretenden reconocer qué valores pueden estar mal calculados y les da un valor acorde a su situación en el mapa.

3.6.1. Media

Asumiendo que un valor de disparidad que difiere mucho del resto de vecinos en una ventana $M \times N$, es un valor erróneo, con este filtro se pretende suavizar estos valores. Como cada valor se verá influenciado por el de sus vecinos, si un valor de disparidad difiere mucho del resto, debe modificarse convenientemente; en cambio, si el valor se asemeja a su vecindad, aquél apenas sufrirá modificación.

Los pasos a seguir para realizar el filtro de la media son:

- 1) Por cada valor $D(a, b)$ del mapa de disparidad se abre una ventana de vecindad de unas determinadas dimensiones ($M \times N$).
- 2) Al valor para el cual se ha abierto la ventana de vecindad se le asigna el valor medio de toda la ventana, ecuación (3.18).

$$D(a, b) = \frac{1}{M \cdot N} \cdot \sum_{i=a-\frac{M}{2}}^{a+\frac{M}{2}-1} \sum_{j=b-\frac{N}{2}}^{b+\frac{N}{2}-1} (D(i, j)) \quad (3.18)$$

Existen dos formas de aplicar este método. Una de ellas tiene en cuenta el valor de los vecinos modificados anteriormente por la media, y en la otra, únicamente se consideran los vecinos del mapa de disparidad original.

- 1) *Teniendo en cuenta los vecinos modificados*: se utiliza el mismo mapa de disparidad (D) tanto para realizar el cálculo como para realizar la modificación, ecuación (3.18).
- 2) *Teniendo en cuenta los vecinos de la disparidad original*: se crea un nuevo mapa de disparidad (ND) igual al mapa original (D), donde las modificaciones se harán sobre el nuevo mapa, y para el cálculo se utiliza el original, ecuación (3.19), de esta forma el valor no se verá influenciado por la modificación de sus vecinos.

$$ND(a,b) = \frac{1}{M \cdot N} \cdot \sum_{i=a-\frac{M}{2}}^{a+\frac{M}{2}} \sum_{j=b-\frac{N}{2}}^{b+\frac{N}{2}} (D(i,j)) \quad (3.19)$$

Una de las ventajas de este filtrado es su sencillez y rapidez. Como se verá en el siguiente capítulo, computacionalmente hablando este filtro es el más rápido de todos, pero en cambio los resultados que obtiene no son precisamente los mejores, ya que modifica, en mayor o menor medida, todos los valores de disparidad obtenidos. Teóricamente, al utilizar la segunda forma del algoritmo donde se tienen en cuenta los vecinos previamente modificados, se debe mejorar el porcentaje de error, ya que utiliza la información previa para generar la nueva disparidad; si bien, como se verá en el capítulo cuatro, no mejora el porcentaje de error aunque sí lo hace su error cuadrático medio.

3.6.2. Mediana

Como en el filtro de la media, se asume que un valor de disparidad que difiere mucho del resto de vecinos en una ventana $M \times N$, es un valor también erróneo, por lo que se pretende suavizar dichos valores erróneos.

Los pasos a seguir para realizar el filtro de la mediana son:

- 1) Por cada valor $D(a,b)$ del mapa de disparidad se abre una ventana de vecindad de unas determinadas dimensiones ($M \times N$).
- 2) Al valor para el cual se ha abierto la ventana de vecindad se le asigna el valor correspondiente a la mediana de toda la ventana, ecuación (3.20).

$$D(a,b) = \text{mediana} \left(D \left[a - \frac{M}{2}, \dots, a + \frac{M}{2}; b - \frac{N}{2}, \dots, b + \frac{N}{2} \right] \right) \quad (3.20)$$

Como con el filtro de la media, existen dos formas de aplicar este algoritmo. En una se tiene en cuenta el valor de los vecinos modificados anteriormente, y en la otra, únicamente se tienen en cuenta los vecinos del mapa de disparidad original.

- 1) *Teniendo en cuenta los vecinos modificados*: Se utiliza el mismo mapa de disparidad (D) tanto para realizar el cálculo como para realizar la modificación, ecuación (3.20).
- 2) *Teniendo en cuenta los vecinos de la disparidad original*: Se crea un nuevo mapa de disparidad (ND) igual al mapa original (D), donde las modificaciones se harán en el nuevo mapa, y para el cálculo se utiliza el original, ecuación (3.21), de esta forma el valor no se verá influenciado por la modificación de sus vecinos.

$$ND(a,b) = \text{mediana} \left(D \left[a - \frac{M}{2}, \dots, a + \frac{M}{2}; b - \frac{N}{2}, \dots, b + \frac{N}{2} \right] \right) \quad (3.21)$$

También, al igual que en el filtro de la media, se trata de un filtro muy sencillo, aunque a diferencia de aquel, su tiempo de procesamiento aumenta debido a la ordenación de los valores de la ventana a la hora de procesar la mediana. Del mismo modo que el caso anterior, el utilizar la segunda forma del algoritmo donde los vecinos son tenidos en cuenta, debería mejorar el resultado. Como se verá también en el capítulo cuatro, en este caso mejora tanto el porcentaje de error como el error cuadrático medio, ya que aunque el filtro se aplica a todos los píxeles y su valor pueda verse modificado, al tratarse de valores enteros, se le asignará un número entero, y no uno real como ocurre en el caso del filtro de la media.

3.6.3. Enfriamiento simulado (Simulated Annealing)

Como en los filtros anteriores, se asume que un valor de disparidad que difiere mucho del resto de vecinos en una ventana $M \times N$, sigue siendo un valor erróneo, por lo que el objetivo sigue siendo tratar de modificar estos valores. Este proceso utiliza una función de energía para la imagen, ecuación (3.26), de esta manera se quiere minimizar este valor en cada paso del algoritmo. Es a esta minimización lo que se le llama Enfriamiento Simulado o “*Simulated Annealing*”. Esta técnica ha sido utilizada por Pajares y Cruz (2004) para correspondencia en visión estereoscópica, si bien para segmentos de borde, por lo que de ella se toma simplemente la idea y la propuesta del algoritmo mostrado en la Figura 3.6, que a su vez se encuentra descrito en Duda y col. (2001).

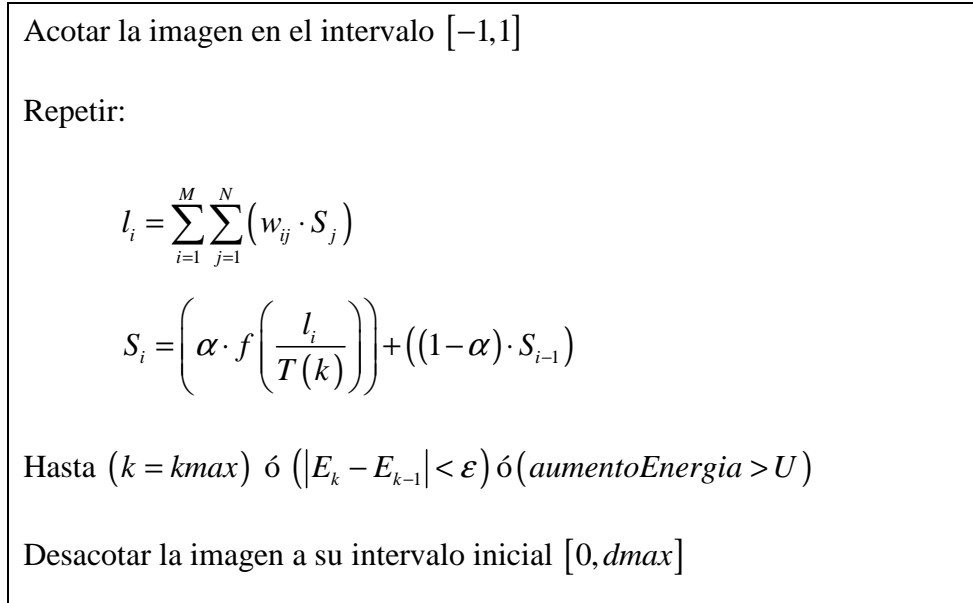


Figura 3.6 - Esquema del algoritmo de enfriamiento simulado

Los detalles del algoritmo mostrado en la figura 3.6 son los siguientes:

- 1) Como entrada recibe el mapa de disparidad previamente calculado.
- 2) Por cada píxel se genera un nodo, y ese nodo tendrá un valor de estado S_i correspondiente al valor de la disparidad en ese momento asociada al píxel que le corresponde.
- 3) Los valores S_i deben estar acotados en el intervalo $[-1,1]$ por lo que antes de aplicar el algoritmo, es necesario tratar la imagen. Para ello, se transforman los

valores de los nodos mediante la ecuación 3.22, que proyecta el rango $[0, dmax]$ al rango $[-1, 1]$. Siendo $dmax$ el valor máximo de disparidad, que se corresponde con el valor máximo de los estados de los nodos de la imagen.

$$S_i = 2 \cdot \frac{S_i}{dmax} - 1 \quad (3.22)$$

- 4) Para la evaluación de los pesos w_{ij} se abre una ventana de vecindad $M \times N$ alrededor del nodo a tratar. Pese a que la imagen se trata como una red neuronal donde todos los nodos están interconectados, únicamente influyen los nodos que pertenecen a la ventana de vecindad N_i^m , por lo que para el resto de nodos que no pertenecen, se anula su peso.

$S_j = 22$	22	22
22	$S_i = 65$	65
21	21	21

Figura 3.7 - Peso entre dos nodos

$$w_{ij} = \begin{cases} 1 - |S_i^{k-1} - S_j^{k-1}| & \forall j \in N_i^m \\ 0 & \forall j \notin N_i^m \end{cases} \quad (3.23)$$

La ecuación (3.23), tiene en cuenta la diferencia entre los valores de los estados de dos nodos (S_i y S_j), otorgando un valor mayor cuanto más similares sean, asignando 1 en caso de ser exactamente iguales, y -1 en caso contrario, es decir, cuando uno tome el valor -1 y el otro 1, o viceversa.

- 5) La función de temperatura T debe ser una función decreciente (ecuación 3.24), en este caso depende del parámetro K que será una constante y representa la temperatura inicial. Esta constante puede o no suavizar más la curvatura, a medida que se aumenta K , disminuyen los valores y por lo tanto se “enfía” más la red. Además, la temperatura disminuye con las iteraciones k

$$T(k) = \frac{K}{\log(k+2)} \quad (3.24)$$

- 6) La función f (ecuación 3.25) es la encargada de modificar el valor obtenido de dividir el sumatorio de los pesos de sus vecinos entre la temperatura actual de la iteración. Esta función acota en el intervalo $[-1.313, 1.313]$ y utiliza como base la tangente hiperbólica (\tanh). Este valor viene determinado por la necesidad de que aumente/disminuya el valor de la operación de división anterior. Tras varias pruebas y ensayos, la inversa de la tangente hiperbólica de 1 es el valor que mejores resultados ha dado (tal y como se puede observar en el capítulo cuatro), haciendo que los valores más bajos disminuyan su valor inicial, y que los valores más altos aumenten, dejando fijos los valores para los cuales se acotó la imagen en el punto 2. De esta forma, los valores de 0, -1 y 1 mantienen su valor, modificándose únicamente los valores comprendidos entre $[-1, 0]$ y $[0, 1]$.

$$f(i) = \frac{\tanh(x)}{\tanh(1)} = 1.313 \cdot \tanh(x) \quad (3.25)$$

- 7) La función de energía (ecuación 3.26) se emplea como criterio de parada del algoritmo, que bajo el enfoque del enfriamiento simulado debe ser decreciente ya además mantenerse por debajo de un cierto valor umbral, U que ha sido fijado a 2 en nuestros experimentos. Para todos los nodos i de la red, se calcula la energía de la ventana de vecindad, la cual se va acumulando para obtener la energía global de la imagen. Pese a que la función deber disminuir, en algún momento puede aumentar para salir de un mínimo global, pero en caso de que aumente en sucesivas iteraciones, o que al disminuir lo haga por debajo de un rango ε , se detiene el algoritmo. De esta forma se pretende ahorrar tiempo y coste computacional.

$$E_i = E_i + \left[-\frac{1}{2} \sum_{i=1}^M \sum_{j=1}^N (w_{ij} \cdot S_j) \right] \quad (3.26)$$

- 8) La proporción α utilizada entre los vecinos y el propio nodo, indica en qué medida los vecinos del nodo influyen en la decisión final. Este valor pertenece al intervalo $[0, 1]$, y como lo que se quiere es suavizar, para ello se tiene que asemejar más a sus vecinos, por lo que dicho valor debe ser superior a 0.5, valor para el cual los vecinos y el nodo tienen la misma influencia. El valor que mejores resultados ha obtenido según los experimentos mostrados en el capítulo

4, es 0.8, ya que tiene en cuenta en su mayoría a los vecinos, y también, aunque en menor medida, el propio valor del nodo.

- 9) Además de utilizar la energía como criterio de parada, se utiliza una variable k . Esta variable se utiliza para realizar un número máximo de iteraciones $kmax$, de esta forma se controla que la evolución de la red no sea excesiva.
- 10) Una vez se ha procesado la imagen, hay que devolver los valores S_i pertenecientes al intervalo $[-1,1]$ a su rango original de disparidades. Para ello, se transforman los valores de los nodos mediante la ecuación (3.27), que desacota del rango $[-1,1]$ al rango $[0,dmax]$. Siendo $dmax$ el mismo valor máximo de disparidad correspondiente al valor máximo de los estados de los nodos de la imagen inicial.

$$d_i = \frac{S_i + 1}{2} \cdot dmax \quad (3.27)$$

3.6.4. Propagación

Tras el estudio realizado en el capítulo cuatro, este filtro se ha creado para ser utilizado junto a la propuesta estudiada en Klaus y col. (2006) e implementada por Lankton (2010), por lo que representa realmente un aporte a la investigación mejorando los resultados obtenidos con la técnica de correspondencia basada en la segmentación. Se ha utilizado este algoritmo de correlación como base, porque obtiene buenos resultados en un tiempo reducido, ya que lo que se busca es que esta técnica pueda ser aplicada en un sistema de tiempo real.

Se trata de un filtro que tiene en cuenta únicamente aquellos píxeles para los cuales no se ha obtenido una buena disparidad mediante la correspondencia basada en la segmentación. A dichos píxeles se les aplica este filtrado, el cual se encarga de propagar los valores de los vecinos para determinar la disparidad correspondiente a los píxeles que mediante la correspondencia no se ha podido definir su valor. Esto hace que el filtrado sea sencillo, rápido y eficiente, además de poder ser implementado en sistemas reales mejorando los tiempos de simulación obtenidos en el siguiente capítulo.

En vista de los resultados obtenidos en el capítulo cuatro, este algoritmo utiliza como base la imagen obtenida al aplicar la correspondencia local basada en la ventana de vecindad, donde todavía no se ha extraído el conjunto de planos de disparidad (paso 3, Figura 3.5).

Partiendo de dicha imagen, los valores para los cuales no se ha podido definir un valor de disparidad aparecen con el valor infinito (∞). Únicamente se van a modificar los valores de dichos píxeles, al contrario que en los filtros anteriores donde todos los son tratados por igual y pueden ver modificado su valor. Para determinar el valor de disparidad, se abre una ventana de vecindad $M \times N$ alrededor del píxel cuyo valor sea infinito, y para el conjunto de valores de vecindad que sí tengan una disparidad fija, se obtiene la mediana (ecuación 3.28), la cual será asignada al píxel de estudio.

$$Disp(a,b) = mediana\left(D\left[a - \frac{M}{2}, \dots, a + \frac{M}{2}; b - \frac{N}{2}, \dots, b + \frac{N}{2}\right]\right) \begin{matrix} \forall Disp(a,b) = \infty \\ \forall D(a,b) \neq \infty \end{matrix} \quad (3.28)$$

Para realizar el recorrido de la imagen, en un primer momento se realizó de izquierda a derecha y de arriba hacia abajo (Dirección 1 de la Figura 3.8). El problema de este recorrido es que la mayoría de los valores de infinito vienen provocados por la oclusión que aparece en las imágenes del par estereoscópico, por lo que no son tenidas en cuenta. Al tratarse de oclusiones hay que considerar que un valor erróneo puede provocar un fallo irreversible en el sistema (por ejemplo, un robot autónomo), y por lo tanto, si no se sabe determinar el valor exacto, conviene dar un valor alto para impedir que dicha zona sea accesible. Por ello, en el algoritmo original se asigna un valor de infinito, si bien lo que idealmente se desea es poder aproximar un valor que se asemeje más a la realidad. Por ello, no se realiza un único recorrido de la imagen en la dirección indicada, sino que se realiza un recorrido en 4 direcciones, partiendo de cada una de las cuatro esquinas de la imagen (Figura 3.8).

Una vez se ha realizado la propagación de los valores en las cuatro direcciones se obtienen cuatro mapas de disparidad filtrados, realizando una nueva pasada sobre la imagen que se ha utilizado como base y asignando el valor máximo de los cuatro filtrados anteriores (ecuación 3.29), al cual se le debe restar una unidad de disparidad a todos los píxeles. Esta disminución se debe a que un estudio previo del algoritmo reveló que los valores que se obtienen mediante esta técnica de correlación, aumentan su

disparidad en dicha unidad con respecto a las imágenes de las cuales se ha considerado, tienen la disparidad real.

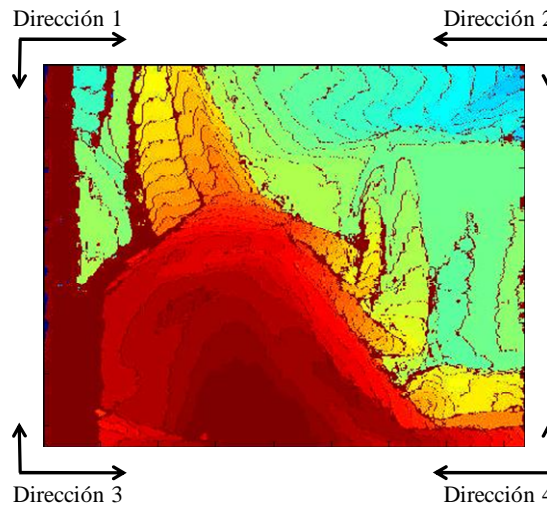


Figura 3.8 - Direcciones del filtrado, siguiendo en cada caso la dirección de la flecha mayor hasta el límite de la imagen y continuando por la dirección de la flecha menor.

$$Disp(a,b) = \max(Dir1(a,b), Dir2(a,b), Dir3(a,b), Dir4(a,b)) - 1 \quad (3.29)$$

A continuación se detallan los pasos del algoritmo propuesto:

1. Obtener la imagen generada al aplicar la correspondencia local basada en la ventana de vecindad, donde todavía no se ha extraído el conjunto de planos de disparidad de la propuesta de Klaus y col. (2006) (paso 3, Figura 3.5).
2. Realizar un recorrido por cada una de las cuatro direcciones de la Figura 3.8 donde para cada valor de disparidad asignado como infinito se abre una ventana de vecindad alrededor del píxel, y cada valor de la ventana distinto de infinito se añade a un conjunto del cual se obtiene la mediana (ecuación 3.28).
3. Una vez obtenidos los filtrados realizados desde las cuatro direcciones, para cada valor de disparidad no definido en la imagen obtenida en el paso 1, se le asigna el máximo de los valores obtenidos en el paso 2 al que se le resta 1 unidad de disparidad (ecuación 3.29).

Una de las principales ventajas de este filtrado es que puede ser paralelizado, al tratarse de realizar cuatro filtrados para obtener el resultado final, todos estos filtrados en las cuatro direcciones pueden realizarse a la vez, de esta forma se reduce aún más el tiempo de ejecución del proceso de la obtención del mapa de disparidad.

RESULTADOS

4.1. Consideraciones preliminares

En este capítulo, se estudian los distintos algoritmos que se han visto en el capítulo anterior. Para ello, se realiza un análisis de los mismos de manera individual para concluir con una visión general de los resultados que se obtienen al aplicar las diferentes técnicas y el filtrado de las imágenes para la mejora del mapa de disparidad.

Los métodos propuestos en este estudio han sido evaluados utilizando imágenes del banco de pruebas de Middlebury (2010) que proporcionan Scharstein y Szeliski (2002). En concreto se han elegido diez pares de imágenes de dicha base de datos cuyas características esenciales son las que se mencionan a continuación:

- 1) Son imágenes de interior con la iluminación perfectamente controlada y, donde las tonalidades de color aparecen con una clara nitidez.
- 2) No poseen distorsión ni aberraciones debido al sistema óptico.
- 3) No poseen desplazamientos verticales, por lo que la búsqueda de correspondencias se puede realizar considerando la componente horizontal.
- 4) Con las imágenes originales, se incluye dos imágenes que denominan “base de verdad” o “ground truth” correspondiéndose con la disparidad supuestamente “real” de cada una de las imágenes estereoscópicas.

La finalidad de la elección de este tipo de imágenes consiste en verificar el comportamiento de los algoritmos con el fin de aislar los problemas inherentes a los mismos, de tal manera que en el siguiente paso consistente en su aplicación al mundo real, sólo queden aquellos problemas derivados de esta naturaleza, tales como: distintos niveles de intensidad en las dos imágenes como consecuencia de brillos y otros efectos derivados de la iluminación en entornos no estructurados, distorsiones radiales de las lentes o desplazamientos verticales de las cámaras.

A lo largo de este capítulo se muestran diferentes pares de imágenes estereoscópicas representativas del conjunto de los diez pares de imágenes utilizados.

La implementación de los distintos métodos que se utilizan en este trabajo se ha realizado en Matlab (2010) y probados con la versión 2009b bajo un procesador Intel Dual-Core T2310 y Sistema Operativo Windows Vista. Matlab se trata de un lenguaje científico interpretado, debido a este hecho, el coste computacional de los procesos implementados en Matlab es superior al que se obtendría con otros lenguajes de distinta naturaleza. No obstante, el objetivo principal de este trabajo no consiste en estudiar los tiempos de cómputo sino el comportamiento de los métodos propuestos, aunque el hecho de tener una idea sobre los mismos puede resultar de interés ya que da idea de su comportamiento cuando el método se traslade a procesamientos con perspectivas de tiempo real.

Los resultados que se presentan en el siguiente estudio, se han conseguido calculando el promedio de los datos obtenidos del conjunto de diez pares de imágenes elegidas al azar del banco de pruebas de Middlebury. En cada algoritmo, se han calculado los datos de dos formas distintas: utilizando una disparidad máxima fijada a un valor de 70, y utilizando la máxima obtenida a partir de las imágenes base proporcionadas.

En cualquier caso, se tienen en cuenta todos los valores de disparidad en la realización de los cálculos, ya que las imágenes de Middlebury contienen valores de disparidad cero, lo que significa que la disparidad no es conocida. Estos valores se deben o bien a una oclusión, o bien a que el objeto está muy cerca de las cámaras.

En los algoritmos utilizados, también es posible que se obtenga un valor de disparidad cero, por ello, siempre que aparezca este valor, o bien en la imagen obtenida, o bien en la que se utiliza para realizar la comparación, si su valor homólogo en la otra imagen del par estereoscópico no es cero se considera un error, es decir, los valores de disparidad cero sí se tienen en cuenta para el cálculo del error. Con esto se consigue obtener el máximo error posible de disparidad, ya que si se omiten estos valores el error se verá reducido, pero así se evita que en algún caso el error pueda aumentar.

4.2. Análisis de los algoritmos para la obtención de la disparidad

4.2.1. Correlación de Pearson

Se trata de un algoritmo sencillo en cuanto a la implementación se refiere, pero es demasiado lento ya que por cada píxel de una determinada imagen se evalúan d_{max} píxeles en la otra. Esto hace difícil su implementación en un sistema a tiempo real, ya que el tiempo de cómputo juega un papel importante en este tipo de sistemas.

A continuación se muestran dos imágenes representativas de un par estereoscópico de los utilizados en la evaluación de los algoritmos (Figura 4.1 y 4.2), así como los resultados obtenidos de aplicar la correlación con la disparidad máxima obtenida de la imagen base (Figura 4.4) y con disparidad fija a 70 (Figura 4.3).



Figura 4.1 - Imagen izquierda del par estéreo.



Figura 4.2 - Imagen derecha del par estéreo.

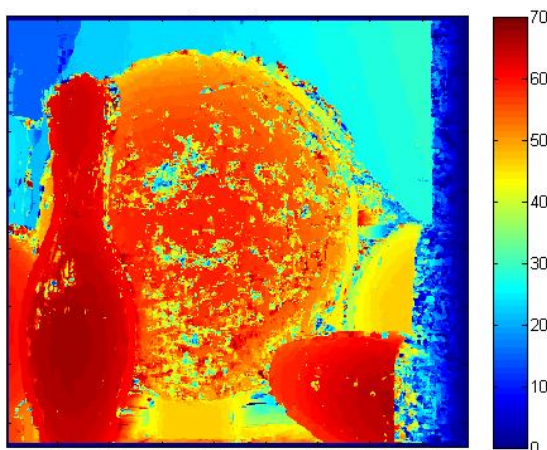


Figura 4.3 - Disparidad obtenida mediante correlación con un valor fijo ($d_{max}=70$).

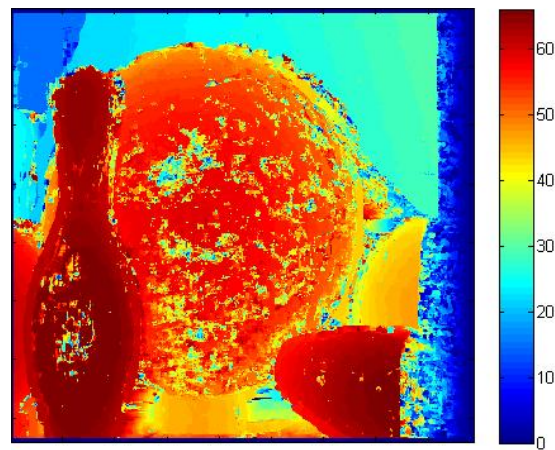


Figura 4.4 - Disparidad obtenida mediante correlación con el valor máximo de la imagen base ($d_{max}=66$).

Como se puede observar a partir de las imágenes anteriores (Figura 4.3 y 4.4), la zona de la bola verde no tiene tanta uniformidad como la obtenida en el bolo blanco. Estos errores son provocados por la obtención de la misma disparidad para distintos píxeles, ya que al utilizar un método basado únicamente en el área y trabajar con una ventana de vecindad, para distintos píxeles dentro del rango d_{max} se puede obtener el mismo valor de disparidad. Además, otro de los problemas de utilizar esta técnica, consiste en que el mismo píxel en ambas imágenes puede variar su valor de intensidad por la cámara utilizada, ya que al ser un componente mecánico, las dos cámaras generalmente no son exactamente iguales.

Dado que las imágenes Figuras 4.3 y 4.4 por sí solas no proporcionan mucha información, para obtener datos relevantes lo que se realiza es una comparación píxel a píxel, representando en negro aquellos píxeles que difieren su valor con respecto a la disparidad esperada, y en blanco, aquellos píxeles para los cuales la disparidad es exactamente la misma, tal y como se puede observar en las Figuras 4.5 y 4.6. Las imágenes mostradas han sido recortadas, ya que se les ha eliminado el margen generado por el algoritmo, así como la zona derecha, debido a que por el propio diseño del algoritmo, para esta zona no se puede hallar una disparidad fiable.

Como este método únicamente tiene en cuenta si un valor es igual a otro, puede ser que al utilizar la media como ecuación de la obtención de la disparidad, dos valores aunque sean próximos, difieran relativamente en una pequeña cantidad. Por ello, se utiliza el error cuadrático medio, obteniendo los siguientes resultados promediados (Tabla 1).

	Tiempo en minutos	Tiempo en segundos	% error imagen sin recortar	Error cuadrático medio imagen sin recortar	% error imagen recortada	Error cuadrático medio imagen recortada
Correlación d_{max}	5,27	316,04	95,38	265,17	94,82	84,67
Correlación fija a 70	6,11	366,61	95,89	269,86	95,40	86,59

Tabla 1 - Resultados medios obtenidos para la ejecución del algoritmo de correlación con disparidad d_{max} y fijada a 70.

Con estos resultados, se puede observar que recortando de la imagen la zona donde la disparidad no es fiable, pese a que el porcentaje de error es alto, para aquellos

píxeles cuyo valor no es el deseado, la variación que sufren con respecto al valor esperado es en promedio de aproximadamente ± 9.5 en ambos casos.

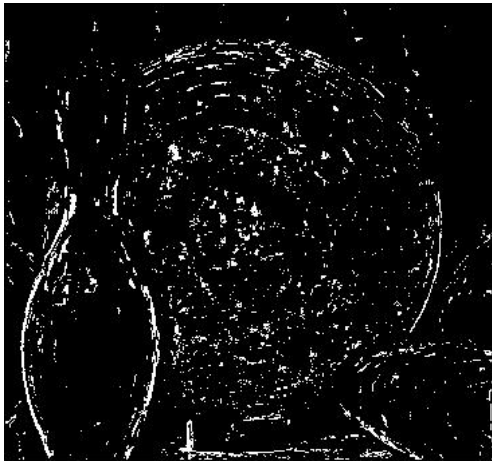


Figura 4.5 - Errores provocados por la correlación con dmax 70 comparada con la imagen base.

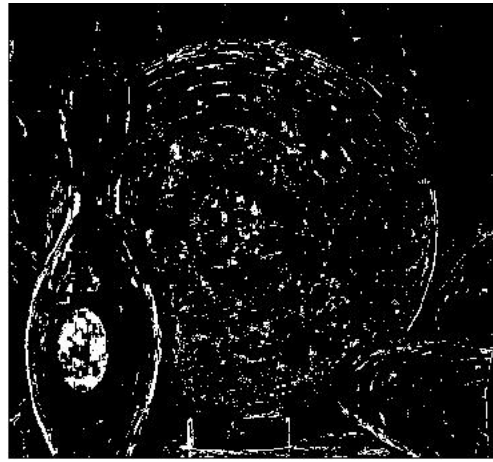


Figura 4.6 - Errores provocados por la correlación con dmax obtenida de la imagen base con la que se compara.

4.2.2. Minimización del error global de energía

Al igual que en el caso de la correlación, no se trata de un algoritmo cuya implementación sea complicada, y aunque mejora el tiempo de ejecución y obtiene un valor de error menor, todavía es demasiado lento por lo que de nuevo su aplicación a sistemas de tiempo real sigue siendo difícil como ocurre con la correlación.

A continuación se muestran dos imágenes originales que representan un par estereoscópico (Figura 4.7 y 4.8), así como los resultados obtenidos tras aplicar el algoritmo de minimización del error de energía global con la disparidad máxima obtenida de la imagen base (Figura 4.10) y con disparidad fija a 70 (Figura 4.9).



Figura 4.7 - Imagen izquierda del par estéreo.



Figura 4.8 - Imagen derecha del par estéreo.

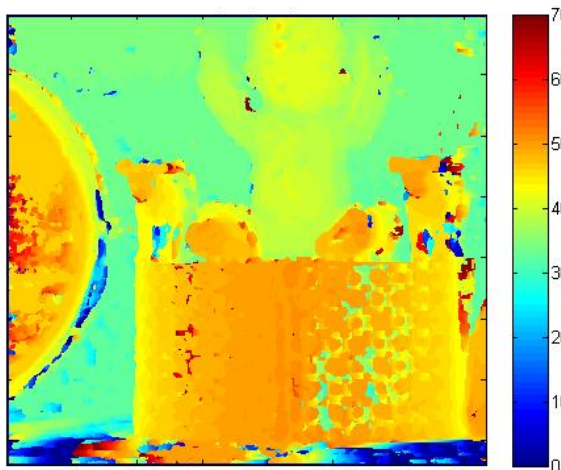


Figura 4.9 - Disparidad obtenida mediante la minimización del error global de energía con un valor fijo ($d_{max}=70$).

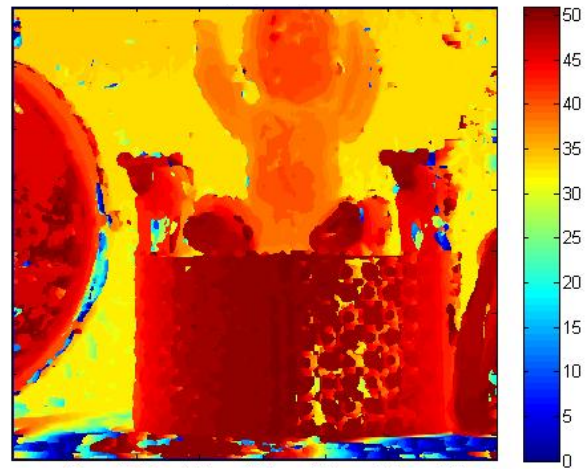


Figura 4.10 - Disparidad obtenida mediante la minimización del error global de energía con el valor máximo de la imagen base ($d_{max}=51$).

De las imágenes anteriores (Figura 4.9 y 4.10) se deduce que ahora los errores son menores que en el caso del algoritmo de correlación previo, ya que los objetos aparecen con una distribución de disparidades mucho más uniforme.

De la misma forma que en el caso anterior se realiza una comparación píxel a píxel, representando en blanco aquellos píxeles para los cuales la disparidad es exactamente la misma y en negro el resto de píxeles que no tienen la misma disparidad, tal y como se puede observar en las Figuras 4.11 y 4.12. En este caso las imágenes son recortadas de igual modo que en el caso anterior.

Para el estudio del error cuadrático medio en este caso, se obtienen los resultados promediados de la Tabla 2.

	Tiempo en minutos	Tiempo en segundos	% error imagen sin recortar	Error cuadrático medio imagen sin recortar	% error imagen recortada	Error cuadrático medio imagen recortada
Correlación d_{max}	3,84	230,29	-	-	41,62	101,67
Correlación fija a 70	4,70	281,95	-	-	42,30	107,48

Tabla 2 - Resultados medios obtenidos para la ejecución del algoritmo de minimización del error global de energía con disparidad d_{max} y fijada a 70.

A partir de los datos obtenidos, se puede observar que para aquellos píxeles cuyo valor no es el deseado, la variación que sufren con respecto al valor esperado es en promedio de aproximadamente ± 10.08 en ambos casos.

Como se puede observar, pese a que ahora el porcentaje de píxeles erróneos es bastante menor en comparación con la correlación, el error cuadrático medio es superior, lo que indica que la variación de cada píxel erróneo es un algo mayor. Al considerarse inapreciable este aumento del error cuadrático medio, en vista de que el tiempo de ejecución es menor y de que el porcentaje de píxeles erróneos es también menor, este algoritmo, en términos generales, puede catalogarse como de mejores prestaciones que el de correlación.

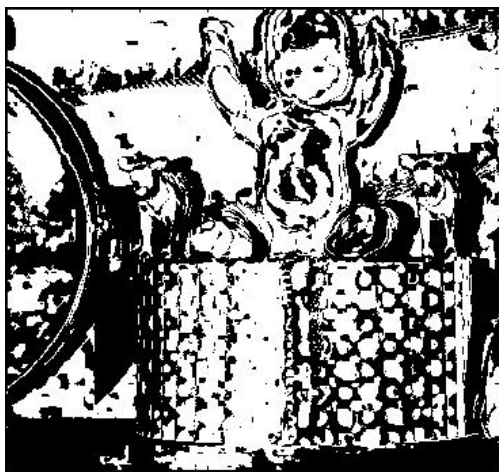


Figura 4.11 - Errores provocados por la minimización del error global de energía con d_{max} 70 comparada con la imagen base.

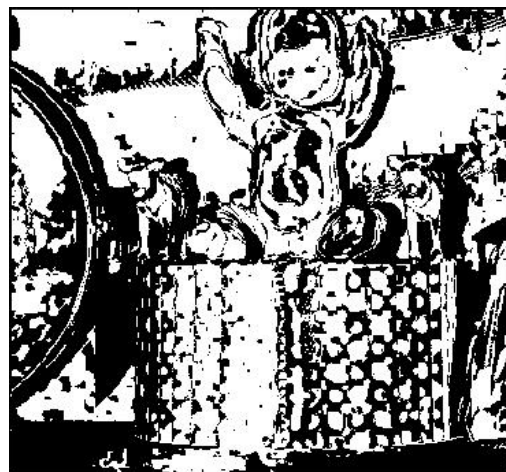


Figura 4.12 - Errores provocados por la minimización del error global de energía con d_{max} obtenida de la imagen base con la que se compara.

4.2.3. Correspondencia estereoscópica basada en línea de crecimiento

Como en los casos previos, no se trata de un algoritmo que plantee dificultades de implementación.

Las imágenes mostradas a continuación representan de nuevo un par de imágenes estereoscópicas originales del conjunto utilizado (Figura 4.13 y 4.14), así como los resultados obtenidos de aplicar el algoritmo de correspondencia basada en línea de crecimiento, tanto con la disparidad máxima obtenida de la imagen base (Figura 4.15) como con disparidad fija a 70 (Figura 4.16).



Figura 4.13 - Imagen izquierda del par estéreo.



Figura 4.14 - Imagen derecha del par estéreo.

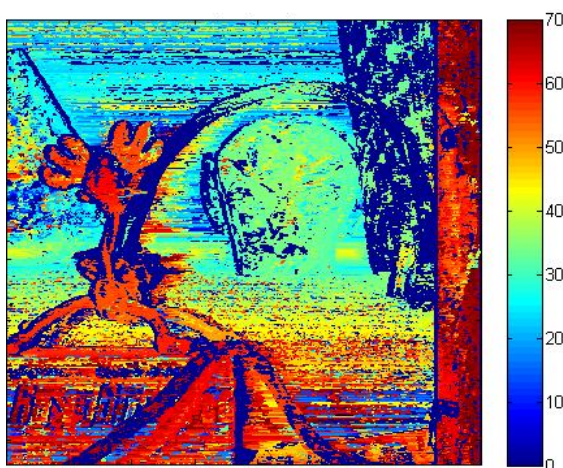


Figura 4.15 - Disparidad obtenida mediante la correspondencia basada en línea de crecimiento con un valor fijo ($d_{max}=70$).

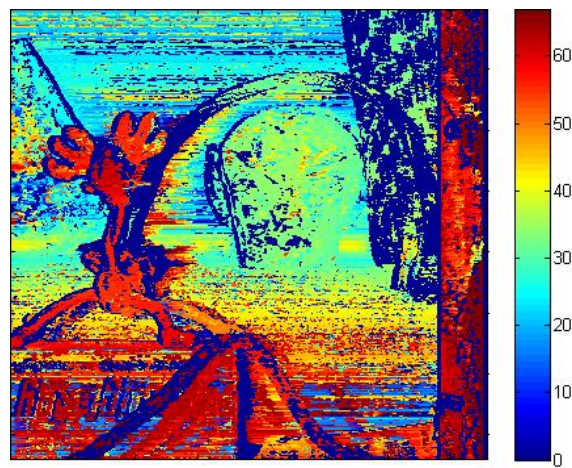


Figura 4.16 - Disparidad obtenida mediante correspondencia basada en línea de crecimiento con el valor máximo de la imagen base ($d_{max}=67$).

Con estas imágenes (Figura 4.15 y 4.16) a simple vista se puede ver que el porcentaje de píxeles erróneos debe ser mayor, ya que las zonas del mismo color apenas presentan uniformidad, y obtiene muchos valores de disparidad cero, los cuales, como ya se ha indicado previamente, casi siempre se tratará de un error.

Como en los casos anteriores se realiza una comparación píxel a píxel, representando los píxeles erróneos y los que han sido correctos, tal y como se puede observar en las Figuras 4.17 y 4.18. En este caso, como en el anterior, las imágenes son recortadas por el propio algoritmo por lo que se les ha quitado el margen.

Para el estudio del error cuadrático medio en este caso, se obtienen los siguientes resultados promediados (Tabla 3).

	Tiempo en minutos	Tiempo en segundos	% error imagen sin recortar	Error cuadrático medio imagen sin recortar	% error imagen recortada	Error cuadrático medio imagen recortada
Correlación dmax	0,20	12,19	-	-	78,49	679,05
Correlación fija a 70	0,20	11,83	-	-	78,51	647,25

Tabla 3 - Resultados medios obtenidos para la ejecución del algoritmo de correspondencia basada en línea de crecimiento con disparidad dmax y fijada a 70

Con estos datos, se observa que para los píxeles erróneos, se obtienen que de media, la variación que sufre con respecto al valor esperado es de aproximadamente ± 26.06 en ambos casos.

Aunque este algoritmo obtenga el mejor resultado en cuanto a tiempo de ejecución, no da buenos resultados en el porcentaje de error y mucho menos en el error cuadrático medio, ya que en promedio por cada píxel erróneo se aproximadamente equivoca en ± 26.06 por lo que no sería viable su implementación en un sistema real.

Por otra parte, aunque mejora el error del algoritmo de correlación, no es tan bueno como el resultado obtenido por la minimización del error global de energía. Pero pese a no obtener el menor porcentaje de error, se trata de un algoritmo relativamente rápido en su ejecución, por lo que resulta ser un candidato para ser implementado en un sistema de tiempo real.



Figura 4.17 - Errores provocados por la correspondencia basada en línea de crecimiento con dmax 70 comparada con la imagen base.



Figura 4.18 - Errores provocados por correspondencia basada en línea de crecimiento con dmax obtenida de la imagen base con la que se compara.

4.2.4. Correspondencia estereoscópica basada en la segmentación

En las Figuras 4.19 y 4.20 se muestran sendas imágenes de un par estereoscópico, así como los resultados obtenidos de aplicar el algoritmo de correspondencia basada en la segmentación utilizando el algoritmo de Lankton (2010), tanto para una disparidad fija a 70 (Figura 4.21) como para la disparidad máxima obtenida de la imagen base (Figura 4.22).



Figura 4.19 - Imagen izquierda del par estéreo.



Figura 4.20 - Imagen derecha del par estéreo.

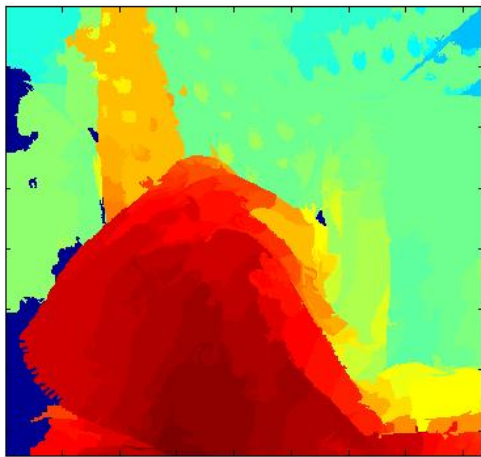


Figura 4.21 - Disparidad obtenida mediante la correspondencia basada en la segmentación con un valor fijo ($d_{max}=70$).

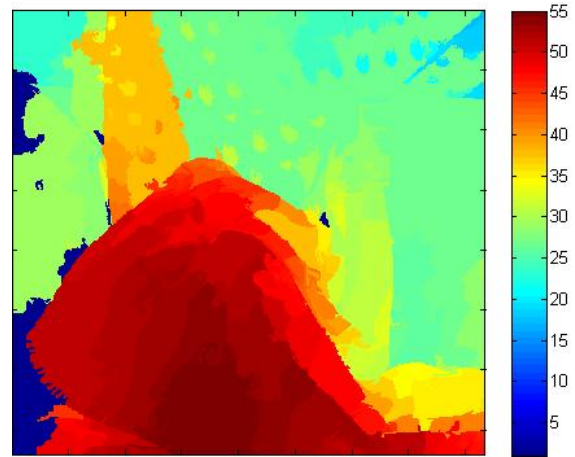


Figura 4.22 - Disparidad obtenida mediante correspondencia basada en la segmentación con el valor máximo de la imagen base ($d_{max}=55$).

De las anteriores imágenes (Figura 4.21 y 4.22) se puede deducir que el porcentaje de error de los píxeles va a ser elevado. Esto se debe a la dinámica del propio algoritmo, ya que asigna un plano de disparidad a un conjunto de píxeles en lugar de asignar un valor independiente a cada punto, lo que ocasiona un grave problema, que se expondrá seguidamente.

Como se puede ver en la Figura 4.24, el resultado que obtiene no es satisfactorio. Esto se debe a la asignación del plano que se realiza, como se asigna el plano dado por la mediada de los píxeles que pertenecen al segmento, esto provoca que objetos esféricos como es la bola o el bolo sean considerados planos.



Figura 4.23 - Imagen real obtenida por la cámara izquierda.

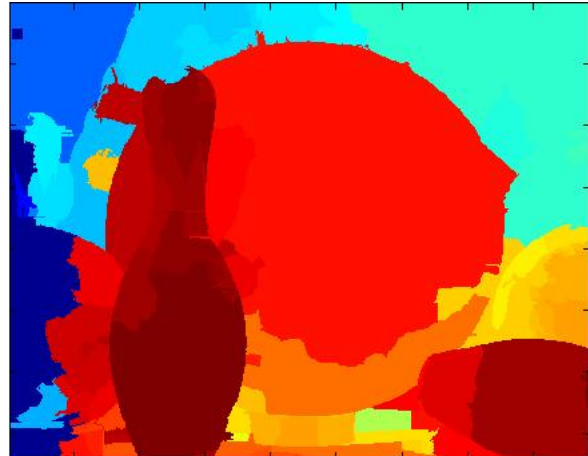


Figura 4.24 - Disparidad obtenida mediante la correspondencia basada en la segmentación.

Debido a este problema dicho algoritmo no podría ser implementado en un sistema real, ya que para este tipo de objetos no se obtiene un valor correcto de distancia sino que se trata de un valor intermedio. Sería conveniente que en lugar de asignar el valor de la mediana de los píxeles que pertenecen al segmento, se asignase el valor máximo, ya que en un sistema real es mejor indicar que un objeto está más próximo de lo que realmente es, a indicar que está más lejos, ya que si se tratase de un robot autónomo ocasionaría una colisión con dicho objeto.

De la misma manera que con el resto de métodos, se realiza una comparación píxel a píxel, representando los píxeles correctos y los que son erróneos, tal y como se puede observar en las Figuras 4.25 y 4.26. En este caso, las imágenes han sido recortadas por el margen izquierdo donde el valor de disparidad obtenido no resulta ser del todo fiable.

La Tabla 4 muestra los el promedio de los valores obtenidos para el estudio del error cuadrático medio de este algoritmo.

	Tiempo en minutos	Tiempo en segundos	% error imagen sin recortar	Error cuadrático medio imagen sin recortar	% error imagen recortada	Error cuadrático medio imagen recortada
Correlación dmax	0,45	27,11	88,09	102,76	89,65	68,22
Correlación fija a 70	0,50	29,69	88,12	102,38	89,86	68,91

Tabla 4 - Resultados medios obtenidos para la ejecución del algoritmo de correspondencia basada en la segmentación con disparidad dmax y fijada a 70

A partir de los datos mostrados en la tabla 4, se observa que para los píxeles erróneos, se obtiene que la variación media que sufre con respecto al valor esperado es de aproximadamente ± 9.1 en ambos casos. Por lo que es el algoritmo que mejor resultados de error cuadrático medio ha dado de todos los estudiados.

Aunque este algoritmo obtenga el mejor resultado en cuanto al error cuadrático medio se refiere, no da buenos resultados en el porcentaje de error, aunque sí tiene un tiempo de ejecución pequeño, por lo que podría ser implementado en un sistema a tiempo real. Por ello, si se modifica este algoritmo como se verá a continuación y además se añade algún filtro que corrija los píxeles cuya disparidad no sea conocida, se puede reducir el porcentaje de error y el error cuadrático medio a costa de aumentar en una pequeña cantidad el tiempo de ejecución.



Figura 4.25 - Errores provocados por la correspondencia basada en la segmentación con dmax 70 comparada con la imagen base.



Figura 4.26 - Errores provocados por correspondencia basada en la segmentación con dmax obtenida de la imagen base con la que se compara.

Un estudio detallado del algoritmo, muestra que la imagen obtenida tras aplicar la correspondencia local basada en la ventana de vecindad no difiere en gran medida de la imagen que se considera que tiene la disparidad real. Para ello se calcula el error cuadrático medio de dicha imagen antes del filtrado de la segmentación junto con la imagen de disparidad real, obteniendo en promedio los resultados mostrados en la Tabla 5:

	% de error imagen sin recortar	Error cuadrático medio imagen sin recortar	% de error imagen recortada	Error cuadrático medio imagen recortada
Correlación dmax	97,24	43,60	96,86	35,21
Correlación fija a 70	97,80	43,68	97,36	35,57

Tabla 5 - Porcentaje de error y error cuadrático medio obtenido de la imagen antes de extraer el conjunto de planos de disparidad.

Con estos datos, se observa que en promedio, por cada píxel se comete un error de ± 6.3 por lo que se trata de un error relativamente reducido. Con el fin de determinar si se puede tratar la imagen para obtener mejores resultados antes de aplicar los filtrados, se obtiene la mediana de las variaciones que sufre cada valor de disparidad obtenido con respecto al valor real. Para ello, debido a que mediante este algoritmo se asigna un valor desconocido “NaN” a aquellos píxeles cuya disparidad no se puede hallar mediante esta correspondencia, únicamente se tienen en cuenta aquellos valores para los cuales se obtiene un valor de disparidad. De este conjunto de valores se calcula la mediana y se observa que en las imágenes de prueba, en su mayoría los errores cometidos son +1, lo que indica que para la mayoría de los píxeles se está cometiendo un exceso de disparidad.

Habiendo visto que en su mayoría la disparidad obtenida está incrementada en uno, se resta una unidad a cada valor de disparidad, y se calculan de nuevo los valores, que se muestran en la Tabla 6.

	% de error imagen sin recortar	Error cuadrático medio imagen sin recortar	% de error imagen recortada	Error cuadrático medio imagen recortada
Correlación dmax	44,91	41,02	39,29	32,21
Correlación fija a 70	44,98	51,86	39,43	32,52

Tabla 6 - Porcentaje de error y error cuadrático medio obtenido de la imagen antes de extraer el conjunto de planos de disparidad modificándola restando uno a cada valor de disparidad obtenido.

De esta forma se ha mejorado aproximadamente en un 53% el porcentaje de error, reduciendo también, aunque en menor medida el error cuadrático medio, siendo ahora ± 6.1 el error medio cometido por cada valor de disparidad.

Como se ha podido observar, en los casos en que las imágenes han sido recortadas se obtienen mejores valores, ya que la zona donde la disparidad no es fiable ha sido eliminada se trata de una zona de incertidumbre, donde el sistema no debería evaluarlo ya que por el diseño del algoritmo no se obtienen resultados correctos para dicha zona. De este modo se reduce tanto el porcentaje de error como el error cuadrático medio.

Por lo tanto, ahora el problema se reduce a calcular valores de disparidad para los píxeles que el algoritmo no es capaz de obtener su valor, además de reducir el porcentaje de error.

4.3. Análisis de los filtros

En este apartado se muestran los resultados de aplicar los diferentes filtrados que se han estudiado, a la imagen que se obtiene del algoritmo de correspondencia basada en la segmentación tras aplicar la correspondencia local y antes de extraer el conjunto de planos de disparidad. Dicha imagen ha sido tratada previamente reduciendo los valores de disparidad obtenidos en una unidad.

4.3.1. Filtro de la media

Este filtrado es de los mejores en cuanto a tiempo se refiere, además de tener una implementación simple. Dicho algoritmo puede ser aplicado el número de veces que se desee, si bien los resultados empeoran cuantas más veces se aplica, ya que los valores de disparidad se ven modificados en gran medida. Esto se debe a que este algoritmo se aplica a todos y cada uno de los valores de disparidad obtenidos, por lo que cuando se transita de un valor que difiere mucho de sus vecinos, se generará una especie de borde difuminado como se puede apreciar en las Figuras 4.27 y 4.28, las cuales son el resultado de aplicar el filtro de la media sin tener en cuenta a los vecinos modificados.

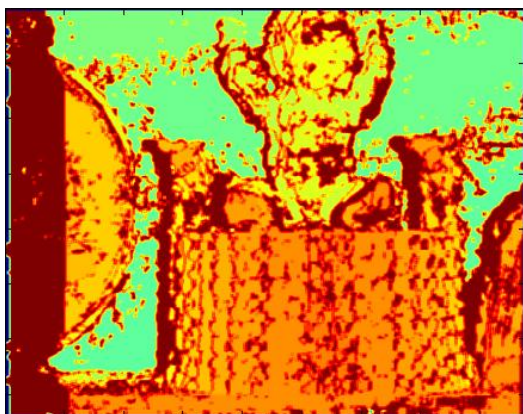


Figura 4.27 - Resultado de aplicar el filtro de la media a una imagen con d_{max} 70.

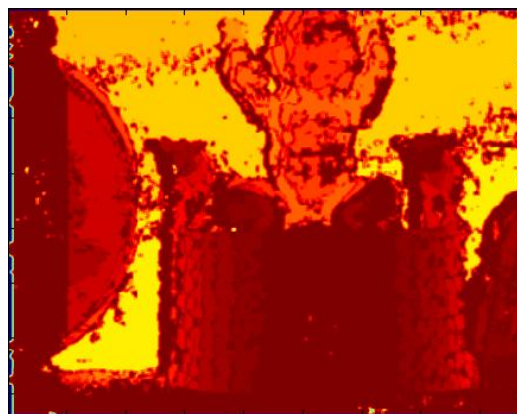


Figura 4.28 - Resultado de aplicar el filtro de la media a una imagen con d_{max} obtenida de la imagen con disparidad real.

Las imágenes 4.29 y 4.30 muestran la comparación píxel a píxel tal y como se ha hecho en los algoritmos anteriores, representando en negro aquellos píxeles para los cuales la disparidad no es igual a la esperada, y en blanco, aquellos píxeles para los cuales la disparidad coincide exactamente con la esperada.



Figura 4.29 - Resultado de aplicar el filtro de la media a una imagen con d_{max} 70.



Figura 4.30 - Resultado de aplicar el filtro de la media a una imagen con d_{max} obtenida de la imagen con disparidad real.

Como se puede inferir a partir de los resultados en la Tabla 7, se trata de un filtro de procesamiento rápido ya que su tiempo de ejecución es muy pequeño, lo que hace que el tiempo de ejecución total apenas se vea incrementado. Si bien incrementa tanto el porcentaje de error como el error cuadrático medio. Esto se debe a que los valores que antes eran enteros, ahora cuando un valor de disparidad difiere del resto dentro de la ventana de vecindad, provoca que dicho valor se modifique y llegue a ser real, ocasionando dicho aumento de errores.

	Tiempo en minutos	Tiempo en segundos	% error imagen sin recortar	Error cuadrático medio imagen sin recortar	% error imagen recortada	Error cuadrático medio imagen recortada	Tiempo minutos algoritmo lankton + filtro	Tiempo segundos algoritmo lankton + filtro
Correlación dmax	0,01	0,44	57,28	213,70	53,31	178,04	0,448	26,758
Correlación fija a 70	0,01	0,38	59,33	330,14	55,32	264,26	0,484	29,028

Tabla 7 - Resultados obtenidos sobre el tiempo, porcentaje de error y error cuadrático medio al aplicar el filtro de la media.

El otro punto de vista del algoritmo consiste en utilizar los valores modificados anteriormente dentro de la ventana de vecindad. Con este enfoque se obtienen los resultados mostrados en las Figuras 4.31 y 4.32.

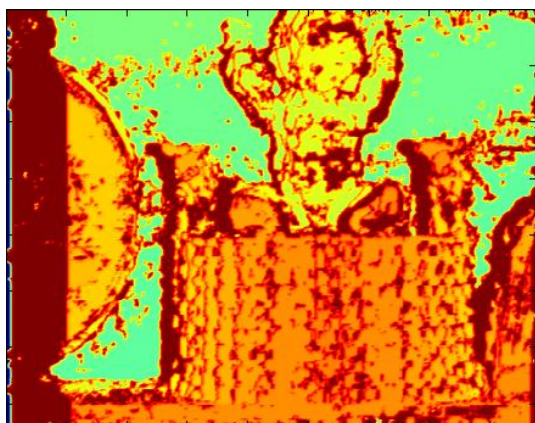


Figura 4.31 - Resultado de aplicar el filtro de la media teniendo en cuenta los vecinos modificados a una imagen con dmax 70.

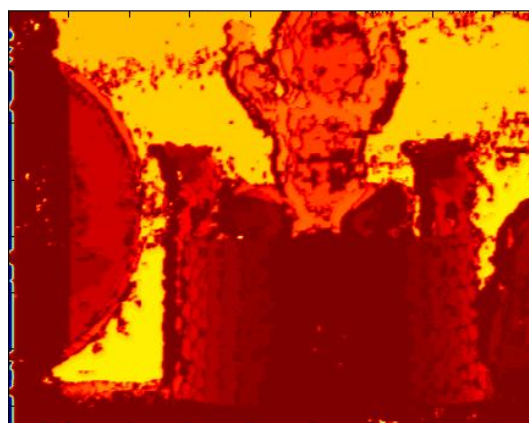


Figura 4.32 - Resultado de aplicar el filtro de la media teniendo en cuenta los vecinos modificados a una imagen con dmax obtenida de la imagen con disparidad real.

Aunque aparentemente las imágenes son similares a las anteriores, en las imágenes de las figuras 4.33 y 4.34 se puede observar cómo el número de píxeles erróneos es distinto. Esto se corrobora con los resultados que se muestran en la Tabla 8. En ella se observa que los valores correspondientes el porcentaje de error han aumentado, mientras que el error cuadrático medio sí ha mejorado. Sin embargo sigue sin mejorar el resultado de la imagen inicial, por lo que tampoco sería válido para ser aplicado en aplicaciones reales.



Figura 4.33 - Resultado de aplicar el filtro de la media a una imagen con dmax 70.

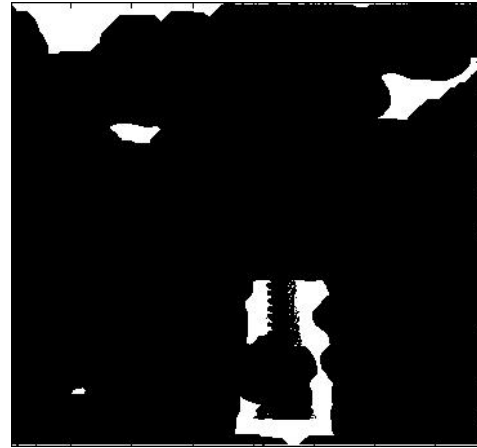


Figura 4.34 - Resultado de aplicar el filtro de la media a una imagen con dmax obtenida de la imagen con disparidad real.

En la Tabla 8 se puede ver que igualmente se trata de un filtro cuya ejecución es relativamente rápida en comparación con los anteriores, por lo que el tiempo de ejecución total apenas se ve incrementado. Si bien, incrementa el porcentaje de error y aunque mejora el error cuadrático medio, tampoco se obtienen unos resultados satisfactorios. Esto se debe a que al tener en cuenta a los vecinos que previamente han sido modificados, aumenta las probabilidades de que el valor sea modificado y pase a tener decimales, por lo que genera un aumento del error.

	Tiempo en minutos	Tiempo en segundos	% error imagen sin recortar	Error cuadrático medio imagen sin recortar	% error imagen recortada	Error cuadrático medio imagen recortada	Tiempo minutos algoritmo lankton + filtro	Tiempo segundos algoritmo lankton + filtro
Correlación dmax	0,01	0,40	97,43	197,89	97,14	162,85	0,448	26,722
Correlación fija a 70	0,01	0,39	98,25	313,85	98,02	247,93	0,484	29,03

Tabla 8 - Resultados obtenidos sobre el tiempo, porcentaje de error y error cuadrático medio al aplicar el filtro de la media.

Para la realización de este estudio, se ha utilizado una ventana de 3x3 para aplicar el filtro de la media en ambos casos. Una ventana mayor provocaría, sobre todo para el caso en el que no se tengan en cuenta los vecinos modificados, un aumento considerable de los errores, debido a que al ser mayor la ventana, las probabilidades de que un píxel difiera del resto también aumentan.

4.3.2. Filtro de la mediana

Dado que el filtro anterior genera un aumento de errores debido al promediado que realiza, se aplica el filtro de la mediana, el cual, pese a modificar los valores, siempre serán enteros, mejorando el porcentaje de error pese a que incrementa el error cuadrático medio.

Como se observa en las Figuras 4.35 y 4.36, el paso de un conjunto de valores que difieren de sus vecinos, ahora no es gradual, por ello, desaparecen los bordes difuminados que aparecían en el filtrado de la media.

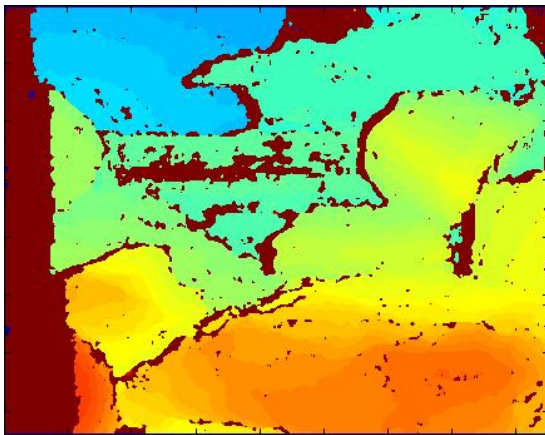


Figura 4.35 - Resultado de aplicar el filtro de la mediana a una imagen con d_{max} 70.

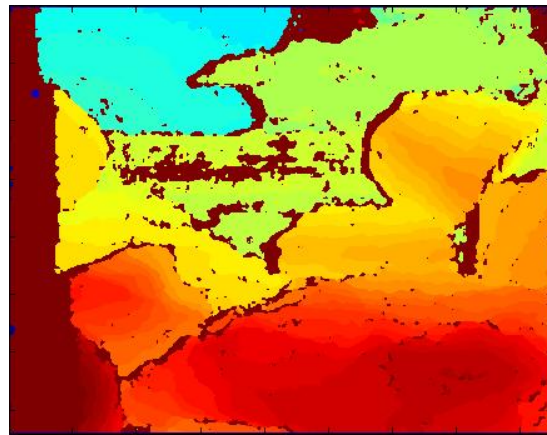


Figura 4.36 - Resultado de aplicar el filtro de la mediana a una imagen con d_{max} obtenida de la imagen con disparidad real.

Las imágenes 4.37 y 4.38 muestran la comparación píxel a píxel tal y como se ha realizado en los algoritmos anteriores, representando en negro aquellos píxeles para los cuales la disparidad no es igual a la esperada, y en blanco, aquellos otros píxeles para los cuales la disparidad resulta ser exactamente la misma.



Figura 4.37 - Resultado de aplicar el filtro de la mediana a una imagen con dmax 70.



Figura 4.38 - Resultado de aplicar el filtro de la mediana a una imagen con dmax obtenida de la imagen con disparidad real.

Este algoritmo presenta un tiempo de ejecución mayor que el de la media, y pese a que mejora el porcentaje de error, no se trata de una mejora significativa, ya que apenas disminuye el porcentaje y aumenta considerablemente el error cuadrático medio tal y como se detalla en la Tabla 9.

	Tiempo en minutos	Tiempo en segundos	% error imagen sin recortar	Error cuadrático medio imagen sin recortar	% error imagen recortada	Error cuadrático medio imagen recortada	Tiempo minutos algoritmo lankton + filtro	Tiempo segundos algoritmo lankton + filtro
Correlación dmax	0,12	7,53	43,37	229,28	37,95	191,30	0,556	33,94
Correlación fija a 70	0,12	7,42	44,33	360,02	38,68	291,61	0,572	36,066

Tabla 9 - Resultados obtenidos sobre el tiempo, porcentaje de error y error cuadrático medio al aplicar el filtro de la mediana.

Considerando el otro enfoque del algoritmo donde se tienen en cuenta los valores modificados anteriormente dentro de la ventana de vecindad, se obtienen los resultados mostrados en las Figuras 4.39 y 4.40.

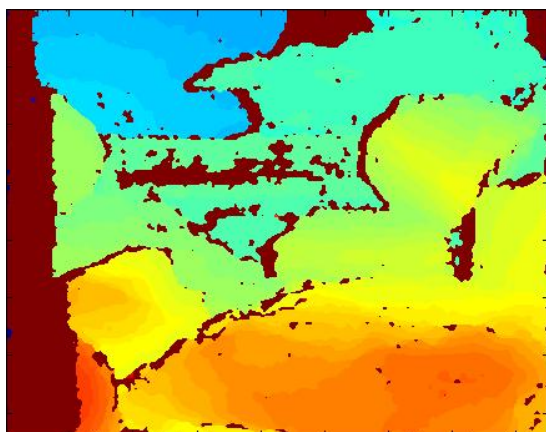


Figura 4.39 - Resultado de aplicar el filtro de la mediana teniendo en cuenta los vecinos modificados a una imagen con d_{max} 70.

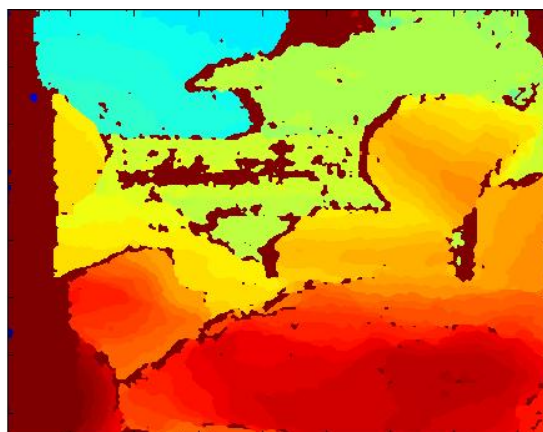


Figura 4.40 - Resultado de aplicar el filtro de la mediana teniendo en cuenta los vecinos modificados a una imagen con d_{max} obtenida de la imagen con disparidad real.

En este caso, se puede apreciar que algunos conjuntos de valores dispersos han desaparecido, y prueba de ello son las Figuras 4.41 y 4.42, donde se muestra gráficamente el porcentaje de error generado. Además, la Tabla 10 recoge los resultados de esta prueba, y se observa que mejora en una pequeña cantidad el porcentaje de error con respecto a la imagen antes del filtro, sin embargo, aunque teniendo en cuenta la vecindad modificada se mejora en comparación a no utilizar dicha vecindad (Tabla 9), el error cuadrático medio ha aumentado en una proporción alta.



Figura 4.41 - Resultado de aplicar el filtro de la mediana a una imagen con d_{max} 70.



Figura 4.42 - Resultado de aplicar el filtro de la mediana a una imagen con d_{max} obtenida de la imagen con disparidad real.

	Tiempo en minutos	Tiempo en segundos	% error imagen sin recortar	Error cuadrático medio imagen sin recortar	% error imagen recortada	Error cuadrático medio imagen recortada	Tiempo minutos algoritmo lankton + filtro	Tiempo segundos algoritmo lankton + filtro
Correlación dmax	0,11	7,25	42,64	211,61	37,15	173,65	0,55	33,50
Correlación fija a 70	0,12	7,50	43,58	339,75	37,83	270,08	0,592	36,142

Tabla 10 - Resultados obtenidos sobre el tiempo, porcentaje de error y error cuadrático medio al aplicar el filtro de la mediana.

Como en el filtro de la media, para la realización de este estudio, se ha utilizado una ventana de 3x3 en ambos casos. Sin embargo, en este caso una ventana mayor no tiene porqué provocar un aumento de los errores, ya que siempre será asignado el valor predominante de toda la ventana, y no se realizaría un simple promediado, como sí ocurre en el caso de la media.

Como este algoritmo apenas mejora el porcentaje de error y además aumenta el error cuadrático medio, tampoco sería válido para un sistema real.

4.3.3. Enfriamiento simulado (Simulated Annealing)

Con el fin de encontrar una mejora de los filtrados anteriores se aplica enfriamiento simulado, el cual pese a promediar los valores lo hace de una forma ponderada, donde no todos los vecinos tienen la misma influencia sobre el punto a evaluar.

Para este estudio los valores que se han utilizado para el número máximo de iteraciones (kmax), ha sido 2, donde los vecinos tienen una influencia de 0.8 y el margen de mejora tiene que ser 0.1, estos valores han sido elegidos ya que tras varias pruebas y ensayos han sido los que mejores resultados han dado.

Las Figuras 4.43 y 4.44 representan el resultado de aplicar el filtro de enfriamiento simulado, en ellas se puede apreciar que la tendencia del algoritmo consiste en llevar todos los valores hacia un mínimo, provocando de gráficamente se vea la imagen como si hubiera sido aclarada.

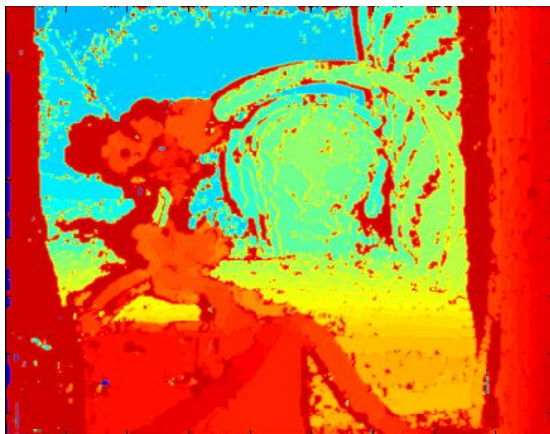


Figura 4.43 - Resultado de aplicar enfriamiento simulado a una imagen con d_{max} 70.

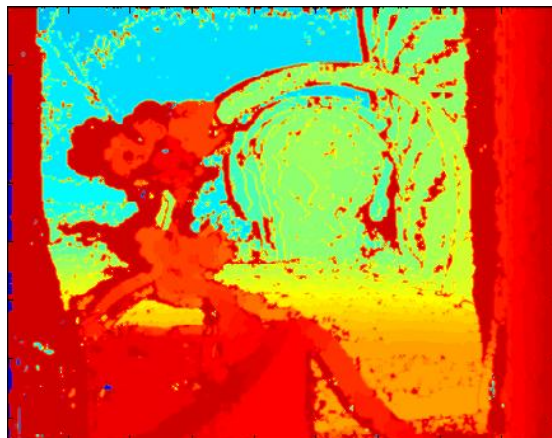


Figura 4.44 - Resultado aplicar enfriamiento simulado a una imagen con d_{max} obtenida de la imagen con disparidad real.

Como en el caso del filtro de la media, en las Figuras 4.45 y 4.46 se observa que el número de píxeles erróneos ha aumentado debido al promediado que se realiza, ya que pese a ser una media ponderada, los valores los convierte en números reales lo que provoca que dichos decimales generen un aumento del porcentaje de error.



Figura 4.45 - Resultado de aplicar enfriamiento simulado a una imagen con d_{max} 70.



Figura 4.46 - Resultado de aplicar enfriamiento simulado a una imagen con d_{max} obtenida de la imagen con disparidad real.

En la Tabla 11 se puede apreciar cómo pese al aumento del porcentaje del error, el error cuadrático medio ha mejorado. Sin embargo, como en el resto de los algoritmos no se mejora los resultados que se obtienen de la imagen que genera la correspondencia basada en la segmentación antes de extraer el conjunto de planos de disparidad, por lo que su implantación en un sistema real queda descartada.

	Tiempo en minutos	Tiempo en segundos	% error imagen sin recortar	Error cuadrático medio imagen sin recortar	% error imagen recortada	Error cuadrático medio imagen recortada	Tiempo minutos algoritmo lankton + filtro	Tiempo segundos algoritmo lankton + filtro
Correlación dmax	0,15	9,18	78,98	148,09	77,52	120,91	0,564	35,106
Correlación fija a 70	0,15	9,18	77,47	223,40	75,90	173,61	0,622	37,752

Tabla 11 - Resultados obtenidos sobre el tiempo, porcentaje de error y error cuadrático medio al aplicar enfriamiento simulado.

4.3.4. Propagación

Dado que con los anteriores algoritmos no se han obtenido resultados satisfactorios, este algoritmo se centra en la propagación de la disparidad de los vecinos de una determinada ventana de vecindad alrededor del punto a evaluar. Los puntos evaluados son únicamente aquellos para los cuales no se tiene una disparidad calculada, con ello se consigue que los píxeles con la disparidad calculada no se vean modificados y únicamente se modifiquen los valores de disparidad considerados como erróneos, consiguiendo una reducción en el porcentaje de error como se verá seguidamente.

Las Figuras 4.47 y 4.48 representan el resultado de aplicar la propagación de las disparidades en las cuatro direcciones tal y como se ha detallado en el capítulo tres.

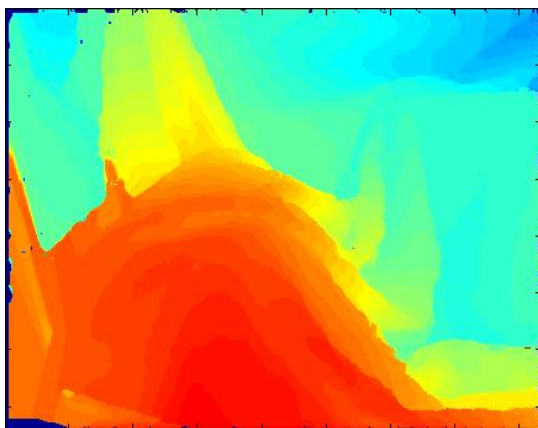


Figura 4.47 - Resultado de aplicar la propagación de disparidad a una imagen con dmax 70.

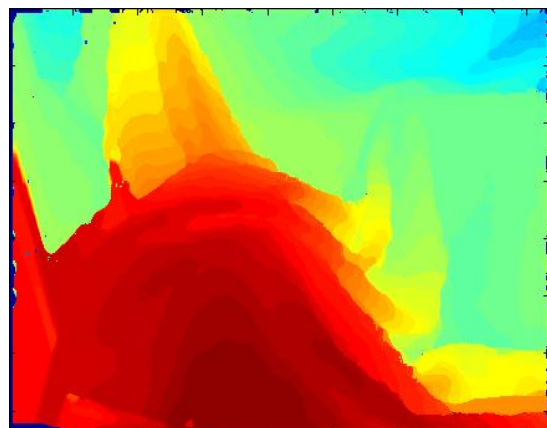


Figura 4.48 - Resultado aplicar la propagación de disparidad a una imagen con dmax obtenida de la imagen con disparidad real.

Con este algoritmo, se puede observar cómo en las Figuras 4.49 y 4.50 el número de píxeles erróneos ha disminuido considerablemente generando errores únicamente en los puntos en los que se pasa entre zonas de distinta disparidad, o lo que es lo mismo entre zonas de diferente color en relación a las imágenes de las Figuras 4.47 y 4.48.



Figura 4.49 - Resultado de aplicar la propagación de disparidad a una imagen con d_{max} 70.

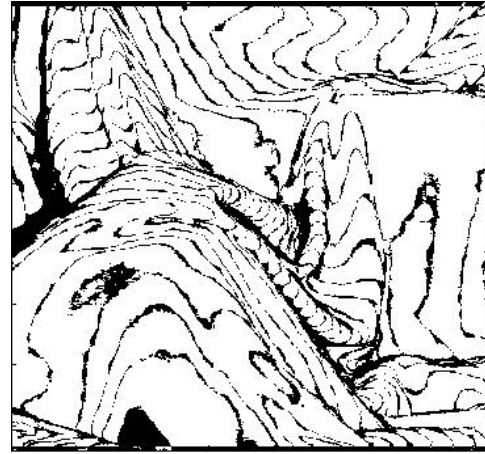


Figura 4.50 - Resultado de aplicar la propagación de disparidad con d_{max} obtenida de la imagen con disparidad real.

De este modo se obtienen los resultados que se pueden apreciar en la Tabla 12, donde el tiempo de ejecución total ha aumentado en unos 27 segundos, consiguiendo reducir un 7% el error cometido. Sin embargo el error cuadrático medio ha aumentado, lo que quiere decir que este algoritmo se equivoca en menos píxeles, pero donde se equivoca el error aumenta en ± 4.5 (pasando de un ± 6.1 a un ± 10.6); si bien, dado que el número de píxeles erróneos es menor, dicho aumento no repercute de manera muy negativa para su implantación en un sistema real.

	Tiempo en minutos	Tiempo en segundos	% error imagen sin recortar	Error cuadrático medio imagen sin recortar	% error imagen recortada	Error cuadrático medio imagen recortada	Tiempo minutos algoritmo lankton + filtro	Tiempo segundos algoritmo lankton + filtro
Correlación d_{max}	0,44	26,97	38,26	119,68	32,91	98,85	0,854	51,468
Correlación fija a 70	0,45	27,20	37,90	117,98	32,58	98,25	0,90	54,174

Tabla 12 - Resultados obtenidos sobre el tiempo, porcentaje de error y error cuadrático medio al aplicar la propagación de disparidad.

Para la realización de este estudio, se ha utilizado una ventana de 7x7 ya que cuanto menor sea el tamaño, menor es la cantidad de valores que se utilizan en la propagación y peores serán los resultados, y cuanto mayor es la ventana, mayor tiempo computacional se necesita haciendo que no compense el tiempo de computo con el resultado obtenido.

4.4. Resumen de los resultados

A continuación se muestra un resumen, Tablas 13 y 14, de los resultados obtenidos por los algoritmos anteriores así como el resultado de aplicar los diferentes filtrados a la imagen obtenida de la correspondencia basada en la segmentación justo después de aplicar la correlación basada en la ventana de vecindad y antes de la extracción del conjunto de planos de disparidad. Además, en las Figuras 4.51 y 4.52 se muestran gráficamente estos mismos resultados para una mejor visualización.

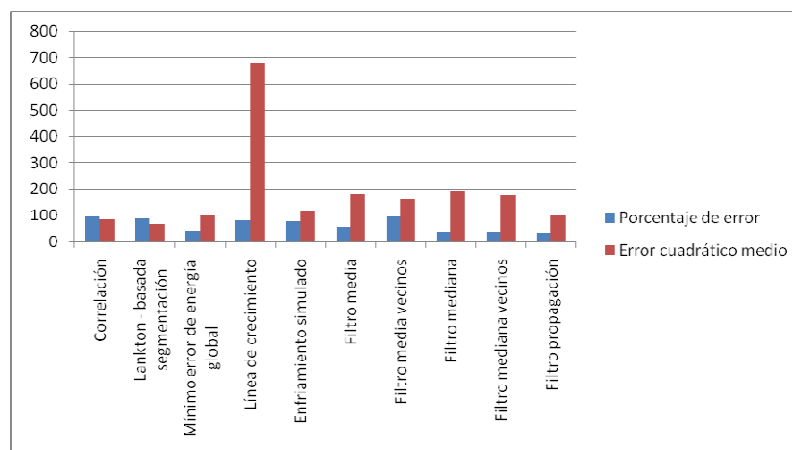


Figura 4.51 - Gráfico que muestra el resultado de los algoritmos y filtro con dmax obtenida de la imagen con disparidad real.

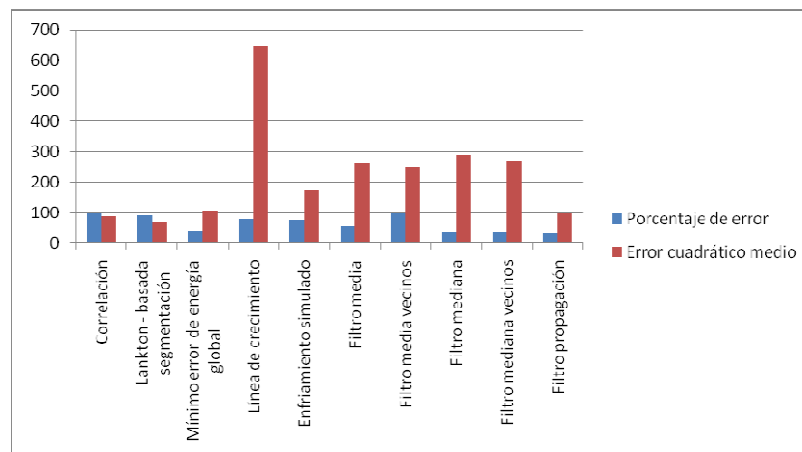


Figura 4.52 - Gráfico que muestra el resultado de los algoritmos y filtro con dmax fijo a 70.

	Tiempo minutos dmax	Tiempo segundos dmax	% error sin recortar	Error cuadrático medio sin recortar	% error recortada	Error cuadrático medio recortada	Tiempo minutos dmax lankton + filtro	Tiempo segundos dmax lankton + filtro
ALGORITMOS ESTEREO								
Correlación	5,27	316,04	95,38	265,17	94,82	84,67	-	-
Lankton - basada segmentación	0,45	27,11	88,09	102,76	89,65	68,22	-	-
Mínimo error de energía global	3,84	230,29	-	-	41,62	101,67	-	-
Línea de crecimiento	0,20	12,19	-	-	78,49	679,05	-	-
FILTRO DE LANKTON ANTES DE EXTRAER EL CONJUNTO DE PLANOS DE DISPARIDAD MODIFICADO								
Enfriamiento simulado	0,15	9,18	78,98	148,09	77,52	120,91	0,564	35,106
Filtro media	0,01	0,44	57,28	213,70	53,31	178,04	0,448	26,758
Filtro media vecinos	0,01	0,40	97,43	197,89	97,14	162,85	0,448	26,722
Filtro mediana	0,12	7,53	43,37	229,28	37,95	191,30	0,556	33,94
Filtro mediana vecinos	0,11	7,25	42,64	211,61	37,15	173,65	0,55	33,50
Filtro propagación	0,44	26,97	38,26	119,68	32,91	98,85	0,854	51,468

Tabla 13 - Promedio de los resultados obtenidos utilizando como disparidad máxima el valor de máximo de la imagen de disparidad real utilizada como base.

	Tiempo minutos disparidad fija a 70	Tiempo segundos disparidad fija a 70	% error sin recortar	Error cuadrático medio sin recortar	% error recortada	Error cuadrático medio recortada	Tiempo minutos fijo a 70 algoritmo lankton + filtro	Tiempo segundos fijo a 70 algoritmo lankton + filtro
ALGORITMOS ESTEREO								
Correlación	6,11	366,61	95,89	269,86	95,40	86,59	-	-
Lankton- basada segmentación	0,50	29,69	88,12	102,38	89,86	68,91	-	-
Mínimo error de energía global	4,70	281,95	-	-	42,3	107,48	-	-
Línea de crecimiento	0,20	11,83	-	-	78,51	647,25	-	-
FILTRO DE LANKTON ANTES DE EXTRAER EL CONJUNTO DE PLANOS DE DISPARIDAD MODIFICADO								
Enfriamiento simulado	0,15	9,18	77,47	223,4	75,90	173,61	0,622	37,752
Filtro media	0,01	0,38	59,33	330,14	55,32	264,26	0,484	29,028
Filtro media vecinos	0,01	0,39	98,25	313,85	98,02	247,93	0,484	29,03
Filtro mediana	0,12	7,42	44,33	360,02	38,68	291,61	0,572	36,066
Filtro mediana vecinos	0,12	7,50	43,58	339,75	37,83	270,08	0,592	36,142
Filtro propagación	0,45	27,20	37,90	117,98	32,58	98,25	0,90	54,174

Tabla 14 - Promedio de los resultados obtenidos utilizando como disparidad máxima un valor fijo de 70.

CONCLUSIONES Y TRABAJO FUTURO

5.1. Conclusiones

A lo largo del presente trabajo se han estudiado diferentes algoritmos de correspondencia estereoscópica con el fin de identificar de entre ellos los más favorables para su aplicación en entornos reales. En principio, este estudio se ha centrado en la evaluación de los mismos desde el punto de vista de su efectividad en términos de rendimiento y en tiempos de cómputo. La evaluación se ha hecho sobre imágenes proporcionadas por la base de datos mencionada dado que además de poseer imágenes fiables, se dispone de mapas de disparidad de referencia.

Según los resultados expuestos a lo largo del capítulo cuatro, el algoritmo que mejores resultados ha generado es el basado en la combinación de la correspondencia mediante la segmentación junto con el filtro denominado propagación de disparidad.

La modificación realizada las imágenes junto con el filtrado, permite reducir aproximadamente alrededor de un 57% el error obtenido, ello pese a que cada error a nivel de píxel que se comete aumenta aproximadamente en un +1.5, cantidad que resulta asumible debido a la considerable mejora del porcentaje de error, ya que aunque el error aumenta, lo hace de forma insignificante al haber menos píxeles erróneos.

Además cumple con la condición de que debe de completarse en un tiempo medianamente reducido, ya que a costa de aumentar el tiempo de ejecución del algoritmo unos segundos más, se obtiene la mejoría indicada. Por tanto, ésta se perfila como una posible solución para su traslado a sistemas reales.

En el estudio realizado se ha mostrado la disparidad obtenida asignando un valor fijo de 70 y un valor variable al ancho de la ventana de disparidad (d_{max}) que se determina en función de la imagen de disparidad real. No obstante, en el futuro cuando se trabaje con sistemas reales, este valor será desconocido por lo que será necesario establecer algún tipo de criterio para fijar dicho valor. Además, como se ha visto en las diferentes imágenes mostradas, existen zonas que por el diseño del algoritmo la

disparidad obtenida no es fiable y por lo tanto se hace necesario acotar dichas zonas en las imágenes, que también dependerá de las condiciones del diseño físico del sistema.

5.2. Trabajo futuro

Uno de los primeros trabajos de futuro que se plantean es el de la implementación de los algoritmos en lenguajes más orientados a tiempo real que Matlab, para tratar de corroborar los resultados obtenidos en este trabajo.

En el traslado de los métodos a la realidad como es el caso de los proyectos mencionados en el capítulo uno, existen diversos problemas, ya identificados que es necesario abordar en futuros trabajos de investigación, en síntesis son los que se enumeran a continuación:

- 1) Las imágenes tomadas de la realidad no suelen ser congruentes en cuanto a niveles de intensidad se refiere, es decir, ambas imágenes deberían tener el mismo nivel de brillo, la misma intensidad e igual contraste, puesto que proceden de cámaras similares, es decir mismo modelo y fabricante, Figuras 5.1 a 5.5. Sin embargo, esto rara vez se consigue. Diversas pruebas en imágenes de interior y de exterior así lo corroboran. En exteriores la problemática es todavía mayor que en interiores. Este problema aparece fundamentalmente cuando la escena no puede ser controlada, siendo totalmente inviable el control de determinadas fuentes de luz, como sin duda es el caso del sol. La distinta posición de las dos cámaras hace que la luz incida de distinta manera sobre cada una de ellas, provocando la aparición de reflejos y diversas reflexiones en los objetos que inciden de diferente manera en sendas imágenes. Se hace pues necesario investigar en este campo para minimizar los efectos mencionados.
- 2) Otro de los problemas que repercuten considerablemente en los resultados, es el debido a la distorsión radial de las lentes y a las aberraciones ópticas. Este fenómeno distorsiona las líneas rectas de la escena real representándolas en sus imágenes como líneas curvas. Por ello, ahora no sólo se centra el problema en la búsqueda de la disparidad, sino que se hace necesaria la investigación sobre los métodos de calibración de las cámaras más apropiados. Este fenómeno queda patente en las imágenes de las Figuras 5.1 y 5.2 tomadas con el robot Surveyor,

Figura 5.3, propiedad del grupo ISCAR (2010). En ellas puede fácilmente apreciarse, los bordes que en el mundo real son rectos, en las imágenes aparecen con una curvatura considerable, aumentando dicha curvatura a medida que se aleja del centro y se acerca a los bordes de la imagen. Esto es lo que técnicamente se conoce como efecto barril.



Figura 5.1 - Imagen izquierda del Surveyor.



Figura 5.2 - Imagen derecha del Surveyor.

- 3) El tercer problema identificado, que requiere de investigación futura, se refiere al hecho de que las cámaras aunque en teoría estén perfectamente alineadas verticalmente, como ocurre en el caso del sistema VIDERE mostrado en la Figura 5.4, siempre se producen desalineamientos verticales, que afectan a la búsqueda de las disparidades, ya que en los sistemas de ejes ópticos paralelos, como son los mostrados en las Figuras 5.3 a 5.5 esto hace que la búsqueda de correspondencias no pueda realizarse sobre las líneas horizontales, esto es las epipolares.

Las Figuras 5.3, 5.4 y 5.5 muestran varios sistemas estereoscópicos diferentes actualmente disponibles en los laboratorios del grupo de investigación ISCAR (2010) sobre los que está previsto desarrollar los trabajos de futuro. En esta línea se pretende revisar conjuntamente los resultados aquí desarrollados con los obtenidos por otros investigadores en el ámbito del grupo ISCAR para continuar la investigación en la línea más apropiada y ya de forma coordinada.

El robot Surveyor, Figura 5.3, es un producto comercial equipado con un sistema *wifi* que permite transmitir las imágenes al computador y desde éste recibir órdenes de desplazamiento. El sistema estereoscópico mostrado en la Figura 5.4 consta de dos cámaras situadas sobre un soporte rígido, con el fin de evitar los mencionados desplazamientos verticales. Además está dotado de diversos orificios con el fin de poder

desplazar las cámaras entre sí, obteniendo distintos ángulos de vista. El rango de variabilidad es de 18 a 60cm. Este sistema se conecta al computador a través del estándar Fire Wire IEEE 1394, proporcionando imágenes en color con resoluciones de hasta 1024x860. Finalmente, en la Figura 5.5 se muestra otro sistema estereoscópico montado sobre un trípode y construido íntegramente en el Laboratorio del Grupo ISCAR con sendas cámaras SCA 140017FC de Basler en color y resoluciones de 1390x1038, fijadas sobre un soporte horizontal. La conexión al computador se realiza también mediante Fire Wire IEEE 1394.

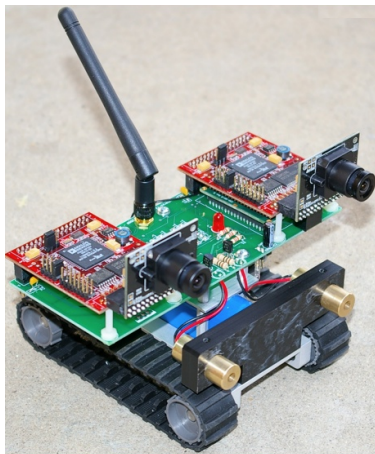


Figura 5.3 - Robot Surveyor.

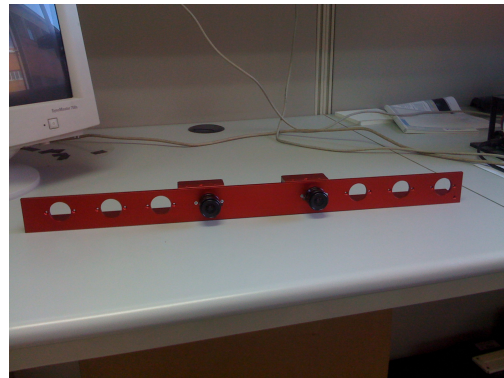


Figura 5.4 - Sistema estereoscópico de línea base variable (distancia horizontal) (VIDERE).



Figura 5.5 - Sistema estereoscópico montado sobre un trípode, con línea base y altura variables

Tal y como se ha expuesto previamente, uno de los problemas identificados es el de la calibración de las cámaras. Sobre este problema existe abundante literatura (Pajares y Cruz, 2007) que es necesario investigar. En este sentido, el objetivo de la calibración es determinar los parámetros que definen el modelo elegido. Este proceso consta de 3 etapas básicas: la primera consiste en la elección del modelo físico, determinando las ecuaciones matemáticas que lo describen, para después proceder a la captura de imágenes proyectando puntos, cuya situación espacial es conocida con elevada precisión sobre el plano sensor, estableciendo así una correspondencia unívoca entre puntos 3D del mundo y puntos 2D sobre el plano sensor. Finalmente se realiza la estimación resolviendo un sistema de ecuaciones por mínimos cuadrados no lineales, tomando como objetivo minimizar el error cuadrático medio.

Para llevar a cabo la calibración de una cámara normalmente se dispone de un objeto patrón, o elemento de referencia, del cual se conocen las coordenadas de algunos de sus puntos (puntos de calibración o de control). Al tomar una imagen, se realiza una proyección de perspectiva que convierte las coordenadas 3D del mundo en coordenadas 2D de la cámara, el objetivo de esta etapa es obtener mediante métodos de Visión Artificial las coordenadas 2D de la cámara y obtener unos ficheros de puntos de control. Después se procede al cálculo de los parámetros intrínsecos y extrínsecos del conjunto cámara, óptica y digitalizador (usando ecuaciones de colinealidad).

El objeto de calibración es una pieza clave a la hora de obtener resultados precisos en la medida, normalmente se suele usar un elemento tridimensional, en el cual destacan una serie de puntos, de los cuales se conoce su posición espacial (puntos de control), al captar el elemento en una imagen, un algoritmo de extracción de características localiza los puntos y los reconoce, estableciendo una asociación entre coordenadas del mundo (x, y, z) y coordenadas de la imagen (x, y) para cada uno de los puntos.

Es importante que el elemento de calibración esté construido con buena precisión. Lo más habitual es esté formado por uno o varios planos sobre los cuales se serigrafían figuras geométricas de forma conocida; como puntos de control se suelen tomar sus vértices o sus baricentros.

En la Figura 5.6 se puede observar un ejemplo de objeto de calibración, donde las cuatro esquinas de cada uno de los cuadrados negros que tiene cada cara son los puntos de calibración usados. También existen técnicas que en lugar de utilizar objetos de calibración 3D utilizan en su lugar mallas 2D colocadas en diferentes orientaciones, para así poder obtener las correspondencias de puntos.

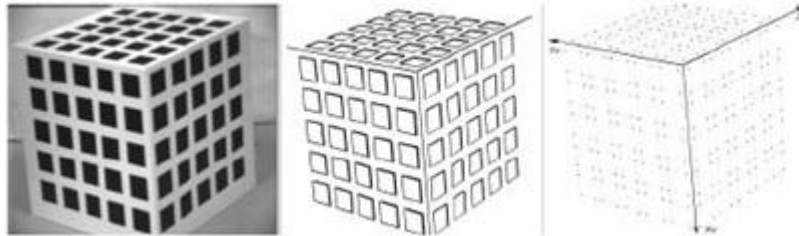


Figura 5.6 - Objetos 3D para la calibración de un sistema estéreo.

BIBLIOGRAFÍA

- Baker, H. Harlyn, Binford, Thomas O. (1981). Depth from Edge and Intensity Based Stereo. *IJCAI 1981*: 631-636
- Barnard, S. (1989). Stochastic stereo matching over scales. *International J. of Computer Vision*.
- Baykant Alagoz, B. (2008). Obtaining Depth Maps From Color Images By Region Based Stereo Matching Algorithms, *OncuBilim Algorithm And System Labs*. Vol. 08, Art.No:04, Computer Vision and Pattern Recognition.
- Bleyer, M., Gelautz, M. (2005). Graph-based surface reconstruction from stereo pairs using image segmentation. In *SPIE*, pages vol. 5665: 288–299, January 2005.
- Chang, C., Chatterjee, S. (1990). Multiresolution stereo - a bayesian approach. *International Conference on Pattern Recognition*.
- Cochran, S.D., Medioni, G. (1992). 3-D Surface Description from binocular stereo, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(10), 981-994.
- Duda, R.O., Hart, P.E. and Stork, D.G. (2001). *Pattern Classification*, Wiley, New York.
- Fatahalian, K., Sugerman, J., Hanrahan, P. (2004). Understanding the Efficiency of GPU Algorithms for Matrix-Matrix Multiplication. Stanford University. Graphics Hardware T. Akenine-Möller, M. McCool (Editors).
- Faugeras, O. (1993). *Three-Dimensional Computer Vision: a Geometric Viewpoint*. The MIT Press, Boston, MA.
- Gamble, E., Geiger, D., Poggio, T., Weinshall, D. (1989). Integration of vision modules and labeling of surface discontinuities. *IEEE Tran. on Systems Man and Cybernetics*.
- Grimson, W.E.L. (1985). Computational experiments with a feature-based stereo algorithm, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7, 17-34.
- Hirschmüller, H., Innocent, Peter R., Garibaldi, Jonathan M. (2002): Real-Time Correlation-Based Stereo Vision with Reduced Border Errors. *International Journal of Computer Vision* 47(1-3): 229-246
- ISCAR (2010). <http://www.dacya.ucm.es/area-isa/index.php?page=home> (disponible on-line)
- Kanade, T., Okutomi, M. (1994). A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment. *IEEE Trans. Pattern Anal. Mach. Intell.* 16(9): 920-932.
- Klaus, A., Sormann, M., Karner, K. (2006). Segmented-Based Stereo Matching Using Belief Propagation and Self-Adapting Dissimilarity Measure, In: *Proc. of 18th Int. Conference on Pattern Recognition*, vol. 3, pp. 15-18.

- Lankton, Shawn M. (2010), (<http://www.shawnlankton.com/2007/12/3d-vision-with-stereo-disparity/>), disponible online)
- Levine, Martin D., O'Handley, Douglas A., Yagi Gary M. (1973). Computer determination of depth maps, *Computer Graphics and Image Processing*, Volume 2, Issue 2, October 1973, Pages 131-150, ISSN 0146-664X, DOI: 10.1016/0146-664X(73)90024-5.
- Lew, M.S., Huang, T.S., Wong, K. (1994). Learning and feature selection in stereo matching, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, pp. 869–881.
- Lew, M. S., Wong, K., y Huang, T. S. (1992). Multiscale stereo matching. *International Conference on Pattern Recognition*.
- López Valles, José M., Fernández Caballero, A., Fernández, Miguel A. (2005). Conceptos y Técnicas de Estereovisión por Computador. *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial*, otoño, 35-62.
- Marapane, S.B., Trivedi, M.M. (1989). Region-based stereo analysis for robotic applications, *IEEE Transactions on Systems, Man and Cybernetics* 19(6), 1447-1464.
- Matlab (2010). The Mathworks (<http://www.mathworks.com/products/matlab/>).
- Matthies, L., Kanade, T., Szeliski, R. (1989). Kalman filter-based algorithms for estimating depth from image sequences. *International Journal of Computer Vision* 3(3): 209-238.
- Matthies, L., Szeliski, R. and Kanade, T. (1988). Kalman Filter-based Algorithms for Estimating depth from Image Sequences. *Proc. DARPA Image Understanding Workshop*, Cambridge, MA, pp. 199-213.
- McKinnon, B., Baltes, J. (2004). Practical Region-Based Matching for Stereo Vision, in: 10th International Workshop on Combinatorial Image Analysis (IWCIA'04), Klette, R., Zunic, J., Eds., Springer, LNCS 3322, pp. 726–738.
- Medioni, G., Nevatia, R. (1985). Segment Based Stereo Matching, *Computer Vision, Graphics and Image Processing*, 31, 2-18.
- Middlebury (2010). (<http://vision.middlebury.edu/stereo>, disponible on-line).
- Moravánszky, A. (2004). Dense Matrix Algebra on the GPU. *NovodeXAG*
- Okutomi, M., Kanade, T. (1993). A Multiple-Baseline Stereo. *IEEE Trans. Pattern Anal. Mach. Intell.* 15(4): 353-363.
- Ohta, Y. and Kanade, T. (1985). Stereo by intra- and inter-scanline search. *IEEE Trans. Pattern Anal. Machine Intell.*, 7(2), 139-154.
- Pajares, G., Cruz, J.M. (2007). *Visión por Computador: Imágenes digitales y aplicaciones*, 2ª ed., RA-MA, Madrid.
- Pajares, G., Cruz, J.M. (2006). Fuzzy Cognitive Maps for stereovision matching, *Pattern Recognition*, 39, 2101–2114.
- Pajares, G., Cruz, J.M. (2004). On combining support vector machines and simulated annealing in stereovision matching, *IEEE Trans. Systems Man and Cybernetics, Part B*, 34(4), 1646-1657.

- Pajares, G., Cruz, J.M. and Aranda, J. (1998). Relaxation by Hopfield Network in Stereo image matching, *Pattern Recognition*, 31(5), 561-574
- Ruichek, Y., Postaire, J.G. (1996). A neural matching algorithm for 3-D reconstruction from stereo pairs of linear images, *Pattern Recognition Letters*, 17, 387-398.
- Scaramuzza, D., Cribblez, N., Martinelli, A., Siegwart, R. (2008). Robust Feature Extraction and Matching for Omnidirectional Images, *Field and Service Robotics*, Laugier, C., Siegwart, R., Eds., Springer, Berlin, Germany, Volume 42, pp. 71–81.
- Scharstein, D., Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, *Int. J. Computer Vision*, vol. 47(1-3), pp. 7–42.
- Shirai, Y. (1987). *Three-dimensional Computer Vision*. Springer-Verlag, Berlin.
- Strintzis, M., Malassiotis, S. (1999). Motor schema-based mobile robot navigation. *IEEE Signal Processing Magazine*.
- Tang, L., Wu, C., Chen, Z. (2002). Image dense matching based on region growth with adaptive window, *Pattern Recognition Letters*, vol. 23, pp. 1169–1178.
- Trucco, E., Verri, A. (1998). *Introductory Techniques for 3-D Computer Vision*. Prentice Hall.
- Venkateswar V., Chellappa R. (1995). Hierarchical stereo and motion correspondence using feature groupings. *International Journal of Computer Vision* 15(3): 245-269.
- Wang, C.Y., Sun, H., Yada, S., Rosenfeld, A. (1983). Some experiments in relaxation image matching using corner features. *Pattern Recognition*. 16(2), 167-182.
- Wei, Y., Quan, L. (2004). Region-Based Progressive Stereo Matching, in: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)*, vol. 1, pp. 106-113.